The Proceedings

OF

THE INSTITUTION OF

ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART C

MONOGRAPHS Nos. 212–245

The Institution of Electrical Engineers

FOUNDED 1871

INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1956–1957

President

SIR GORDON RADLEY, K.C.B., C.B.E., Ph.D.(Eng.).

Past-Presidents

SIR JAMES SWINBURNE, Bart., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROFESSOR E. W. MARCHANT, D.Sc.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.B.E., D.S.O., M.C., T.D., D.Sc.(Eng.).

SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROFESSOR E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.).
SIR JOSIAH ECCLES, C.B.E., D.Sc.
SIR GEORGE H. NELSON, Bart.

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
WILLIS JACKSON, D.Sc., D.Phil., Dr.Sc.Tech., F.R.S.

G. S. C. LUCAS, O.B.E.
SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc.

Honorary Treasurer

THE RT. HON. THE VISCOUNT FALMOUTH.

Ordinary Members of Council

PROFESSOR H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
J. BENNETT.
J. A. BROUGHALL, B.Sc.(Eng.).
C. M. COCK.
A. R. COOPER, M.Eng.
B. DONKIN, B.A.
PROFESSOR J. GREIG, M.Sc., Ph.D.
E. M. HICKIN.
F. J. LANE, O.B.E., M.Sc.
D. McDONALD, B.Sc.
C. T. MELLING, C.B.E., M.Sc.Tech.

H. H. MULLENS, B.Sc.
A. H. MUMFORD, O.B.E., B.Sc.(Eng.).
W. F. PARKER.
D. P. SAYERS, B.Sc.
G. L. WATES, J.P.
H. WATSON-JONES, M.Eng.
D. B. WELBOURN, M.A.
H. WEST, M.Sc.
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.

Chairmen and Past-Chairmen of Sections

Measurement and Control:
D. TAYLOR, M.Sc., Ph.D.
*W. BAMFORD, B.Sc.

Radio and Telecommunication:
R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).
*H. STANESBY.

Supply:
P. J. RYLE, B.Sc.(Eng.).
*L. DRUCQUER.

Utilization:
H. J. GIBSON, B.Sc.
*D. B. HOGG, M.B.E., T.D.

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
H. L. HASLEGRAVE, M.A., Ph.D., M.Sc. (Eng.).
*F. R. C. ROBERTS.

Mersey and North Wales Centre:
P. D'E. STOWELL, B.Sc.(Eng.).
*PROFESSOR J. M. MEEK, D.Eng.

North Midland Centre:
W. K. FLEMING.
*F. BARRELL.

North-Eastern Centre:
*J. CHRISTIE.
A. H. KENYON.

North-Western Centre:
T. E. DANIEL, M.Eng.
*G. V. SADLER.

Northern Ireland Centre:
DOUGLAS S. PARRY.
*MAJOR E. N. CUNLIFFE, B.Sc.Tech.

Western Centre:
PROFESSOR G. H. RAWCLIFFE, M.A., D.Sc.
*T. G. DASH, J.P.

Scottish Centre:
PROFESSOR F. M. BRUCE, M.Sc., Ph.D.
*E. WILKINSON, Ph.D., B.Eng.

South Midland Centre:
C. J. O. GARRARD, M.Sc.
*H. S. DAVIDSON, T.D.

Southern Centre:
H. ROBSON, B.Sc.
*L. H. FULLER, B.Sc.(Eng.).

* Past-Chairman.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary
F. JERVIS SMITH, M.I.E.E.

Assistant Secretary
F. C. HARRIS.

Editor-in-Chief
G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

DISCUSSION ON
'RESULTANT REACTIVE POWER OF OVERHEAD LINES'*

Mr. R. A. Hore (*communicated*): Mr. Coventry's results can be generalized without loss of accuracy in the following manner, thereby providing one chart of universal application:

Surge-impedance loading of a line $= \dfrac{kV^2}{Z}$ megavolt-amperes

where kV = Line-to-line voltage, kilovolts.

Z = Surge impedance of the line $= \sqrt{(L/C)}$ ohms.

Using eqn. (1) of his paper, Mr. Coventry's eqn. (2) reduces to

$$Q = Q_L - Q_C = 1 \cdot 725 \times 10^{-3}[i^2 - 1] \quad . \quad . \text{(A)}$$

where i is the per-unit line current (on the surge impedance load base) and Q is the per-unit reactive power per mile.

Neglecting voltage regulation, as the author has done, the numerical value of i is the same as that of the load. The diagram of Fig. A can thus be plotted to a base of per-unit surge-impedance loading and is applicable to all lines. The chart is used thus:

For a $0 \cdot 175 \, \text{in}^2$ $132 \, \text{kV}$ line the surge-impedance load is $(132^2 \times 5 \cdot 48 \times 10^{-6})/0 \cdot 0021$, or $45 \cdot 5 \, \text{MVA}$.

When transmitting $22 \cdot 8 \, \text{MVA}$ (corresponding to a line current of 100 amp) the per-unit loading is $0 \cdot 502$.

The reactive requirement is $0 \cdot 13 \times 10^{-2}$ per unit per mile (leading), or $59 \, \text{kVAr}$ per mile, which is the result given by Mr. Coventry's Figs. 1 and 2.

Although the whole analysis, generalized or not, is approximate, it is accurate enough to be very useful for rough assessments.

Mr. A. F. Coventry (*in reply*): I think Mr. Hore's contribution is a valuable development of my thesis, and I am very pleased to have it associated therewith. I do not think, however, that the generalized form should supersede entirely the form in which I have presented it, because I think the latter will be more readily comprehensible to the average engineer and the family of curves are revealing in themselves.

Mr. Hore's method is, of course, not perfectly general in that the chart relates to only one frequency, namely 50 c/s. This could easily be remedied, however, by adding curves for 60 and 40 c/s. Alternatively, a note could be added stating that, for frequencies other than 50 c/s, results should be multiplied by $f/50$ where f is the frequency under consideration.
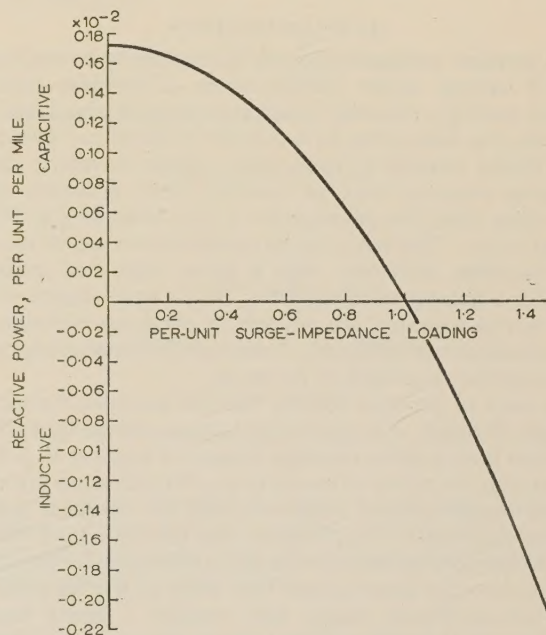

Fig. A.—Resultant reactive power of overhead line.

APPROXIMATE SURGE-IMPEDANCE LOADING

Voltage	Single conductor	Duplex
kV	MVA	MVA
33	2·9	—
66	11·6	—
88	20·6	—
110	32·3	—
132	46·5	—
150	60·0	—
220	129	156
275	202	243
330	290	351
380	—	485

Surge-impedance load $= \dfrac{kV^2 \times 5 \cdot 48 \times 10^{-6}}{L}$

where kV = Line voltage, kilovolts.
L = Line inductance, henrys per mile.

* COVENTRY, A. F.: Monograph No. 172 S, April, 1956 (see **103** C, p. 334).

TRANSIENT RESPONSE OF TWO-TERMINAL NETWORKS

By O. P. D. CUTTERIDGE, M.Sc.(Eng.), Ph.D.

SUMMARY

The maximum possible increase in gain/rise-time for a 2-terminal network with shunt capacitance is shown to be 2·75 compared with the parallel RC case; the overall gain/rise-time figure for a cascade of n such stages is also deduced. As such networks would have a large amount of overshoot, those having a monotonic response to a step-function current drive are next considered. Improvement in figure of merit is shown to be greatest with complex poles and zeros, and the case of three poles, two of which are conjugate complex, is discussed in some detail. Finally, modifications are made to a network with three poles having a monotonic response so as to produce improved rise-time at the expense of a small amount of overshoot; e.g. a figure of merit of 1·90 can be obtained with 1% overshoot.

(1) INTRODUCTION

The transient voltage developed, in response to a step-function drive of current, across various forms of low-pass 2-terminal network having a parasitic capacitance shunted across the input terminals, has been given by a number of authors. Two points of particular interest in the output voltage waveform are the percentage overshoot and the rise-time, which is usually defined as the time taken for the response to rise from 10% to 90% of its final value. The basic design problem is to secure the minimum rise-time consistent with a given degree of overshoot. Since, for a number of applications, only a small degree of overshoot can be tolerated, the problem of obtaining minimum rise-time consistent with monotonic* response is of special importance and forms the main topic of the paper.

The work of previous authors (see, for example, Reference 1) has generally been of the following form: a certain network configuration having given elements connected together in a certain way is taken, the values of one or more of the elements are varied, and the combination of values yielding the optimum transient response is noted. This procedure has the advantage that only forms of network suitable for physical realization are considered, with the possible disadvantage that some of the networks that were not considered might have yielded a better transient response. It is therefore of interest to examine the optimum transient responses possible with stable, passive, 2-terminal networks having a given parasitic shunt capacitance. This is the main topic considered in the following Sections.

(2) MINIMUM POSSIBLE RISE-TIME FOR A TWO-TERMINAL NETWORK

The general expression for a low-pass 2-terminal impedance, $Z(p)$, having a zero frequency value of G and parasitic shunt capacitance of C, can be written in the form

$$Z(p) = G\frac{1 + a_1(pGC) + a_2(pGC)^2 + \ldots + (pGC)^{n-1}}{1 + b_1(pGC) + b_2(pGC)^2 + \ldots + (pGC)^n} \quad . \quad (1)$$

* A monotonic function of t is one which never decreases (or never increases) with increasing t.

---

where p is the complex frequency or the parameter of the Laplace transform. If $f(t)$ is the time domain response to a unit step-function drive and $f_n(t)$ is this response normalized to unity at time infinity, then the p-multiplied Laplace transform[2] of $f_n(t)$ is given by

$$p\int_0^\infty f_n(t)\varepsilon^{-pt}dt = \frac{1 + a_1(pGC) + a_2(pGC)^2 + \ldots + (pGC)^{n-1}}{1 + b_1(pGC) + b_2(pGC)^2 + \ldots + (pGC)^n}$$
$$\ldots \quad (2)$$

But by the power-series expansion theorem[3] this transform can also be written as

$$p\int_0^\infty f_n(t)\varepsilon^{-pt}dt = f_n(0) + \frac{f_n'(0)}{p} + \frac{f_n''(0)}{p^2} + \text{etc.} \quad . \quad (3)$$

Thus $f_n(t)$, plotted against a normalized time scale of t/GC, passes through the origin at $t = 0$ and begins to rise with a slope of unity. Because this finite rate of rise at the origin will invariably carry the response over the 10% mark at which rise-time measurements are commenced, minimum rise-time for a 2-terminal network is obtained by adjusting the time of occurrence of 90% of the final response to be as low as possible. The maximum rate of rise of output voltage, across the parasitic capacitance, will occur if all the step-function driving current proceeds through this capacitance and none enters the rest of the network until 90% of the output voltage is attained. The curve of minimum rise-time for a 2-terminal network, when plotted against the variable t/GC, is thus a straight line of unit slope up to the 90% mark. The best possible gain/rise-time quotient attainable with a 2-terminal network is therefore given by

$$\text{Gain/rise-time} = \frac{1 \cdot 25}{C} \quad . \quad . \quad . \quad . \quad (4)$$

The gain/rise-time quotient for the simplest possible 2-terminal network consisting of a parallel RC network is easily found to be $1/2 \cdot 197C$. It is useful to define the figure of merit of a network as

$$\text{Figure of merit of network} = \frac{\text{Gain/rise-time of network}}{\text{Gain/rise-time of parallel } RC \text{ network}}$$

Hence the maximum possible figure of merit of a 2-terminal network is $2 \cdot 75$. A network could have a figure of merit approaching this value only if a large amount of overshoot could be tolerated. The value for the maximum possible figure of merit for a 2-terminal network having monotonic response cannot be obtained in such a simple manner but would be less than that quoted above.

The maximum possible figure of merit can be obtained with the simple network* of Fig. 1 as $L \to \infty$. The transient response of this network to a unit step of current is given by

$$1 - \varepsilon^{-t/2L}\left[\cos\frac{\sqrt{(4L-1)}}{2L}t - \frac{(2L-1)}{\sqrt{(4L-1)}}\sin\frac{\sqrt{(4L-1)}}{2L}t\right] \cdot (5)$$

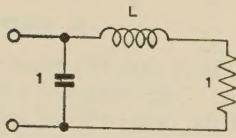* Throughout the paper the units for capacitance, resistance and inductance are farads, ohms and henrys, respectively.

Fig. 1.—Network realizing maximum possible figure of merit as $L \to \infty$.

and it can hence be easily shown that the figure of merit approaches $2 \cdot 75$ as L is made indefinitely large.

For a cascade of n stages, each having this minimum rise-time and separated from each other by valves of mutual conductances $g_{m1}, g_{m2}, \ldots, g_{mn}$, the overall gain/rise-time figure becomes

$$\frac{n\text{th root of overall gain}}{\text{Overall } 10\%\text{–}90\% \text{ rise-time}}$$

$$= \sqrt[n]{\left(\frac{g_{m1}}{C_1} \times \frac{g_{m2}}{C_2} \cdots \times \frac{g_{mn}}{C_n}\right)} \frac{1}{K(n)} \quad . \quad . \quad . \quad (6)$$

where $K(n) = \sqrt[n]{n!}(\sqrt[n]{0 \cdot 9} - \sqrt[n]{0 \cdot 1})$.

Fig. 2 shows the behaviour of the function $K(n)$ for $1 \leqslant n \leqslant 100$; it is seen that the function is practically constant over this range of n.
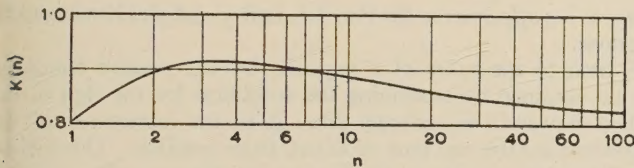

Fig. 2.—Graph of the function $K(n) = \sqrt[n]{n!}(\sqrt[n]{0 \cdot 9} - \sqrt[n]{0 \cdot 1})$.

(3) TWO-TERMINAL NETWORKS HAVING MINIMUM RISE-TIME CONSISTENT WITH MONOTONIC RESPONSE

(3.1) Derivation of Rise-Time from the System Function

Although the 10%–90% rise-time is, in principle, determined when the numerator and denominator coefficients of a system function $Z(p)$ are given, it seems extremely difficult to determine its value directly from the system function. For this reason other definitions of rise-time have been given which yield simple expressions in terms of the system function. Three such definitions have been given by Mathers,[4] but that given by Elmore,[5] and applicable only in the case of monotonic response, will be used here. According to Elmore, the rise-times calculated on his basis generally differ by less than 10% from those calculated by the 10%–90% method. In terms of $f'_n(t)$, the time derivative of the normalized response, Elmore's definition of rise-time, T_R, is given by

$$T_R^2 = 2\pi \int_0^\infty (t - T_D)^2 f'_n(t) dt \quad . \quad . \quad . \quad (7)$$

where

$$T_D = \int_0^\infty t f'_n(t) dt$$

If $F(p)$ is the system function corresponding to a normalized step-response $f_n(t)$, i.e.

$$F(p) = p \int_0^\infty f_n(t) \varepsilon^{-pt} dt \quad . \quad . \quad . \quad . \quad (8)$$

and

$$F(p) = \frac{1 + a_1 p + a_2 p^2 + \ldots + a_m p^m}{1 + b_1 p + b_2 p^2 + \ldots + b_n p^n} \quad . \quad . \quad (9)$$

then T_R is given by

$$\frac{T_R^2}{2\pi} = b_1^2 - a_1^2 + 2(a_2 - b_2) \quad . \quad . \quad . \quad (10)$$

Now $F(p)$ can also be written in the form

$$F(p) = \frac{(1 - p/p_1)(1 - p/p_2) \ldots (1 - p/p_m)}{(1 - p/p'_1)(1 - p/p'_2) \ldots (1 - p/p'_n)}$$

$$= \frac{1 - p\left(\sum_1^m \frac{1}{p_i}\right) + p^2\left(\sum_{\substack{1 \\ i < j}}^m \frac{1}{p_i p_j}\right) + \ldots +}{1 - p\left(\sum_1^n \frac{1}{p'_i}\right) + p^2\left(\sum_{\substack{1 \\ i < j}}^n \frac{1}{p'_i p'_j}\right) + \ldots +} \quad . \quad (11)$$

where $p_1, p_2, \ldots, p_m$ and $p'_1, p'_2, \ldots, p'_n$ are the zeros and poles respectively of $F(p)$. Comparing with the previous expressions it is seen that

$$\frac{T_R^2}{2\pi} = \sum_1^n 1/p_i'^2 - \sum_1^m 1/p_i^2 \quad . \quad . \quad . \quad (12)$$

Thus real poles add to and real zeros subtract from the rise-time. The rise-time contribution from a conjugate complex pair of poles, p'_1 and p'_2, is

$$\frac{1}{p_1'^2} + \frac{1}{p_2'^2} = \frac{1}{|p_1'|^2} 2 \cos 2\phi \quad . \quad . \quad . \quad (13)$$

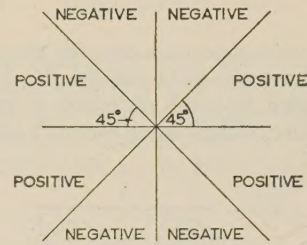where ϕ is the phase angle of p'_1.

Fig. 3.—Sign of the resultant contribution from a pair of conjugate complex poles.

Fig. 3 shows the sign of the resultant contribution from a pair of conjugate complex poles according to their position on the complex plane; the signs for a pair of conjugate complex zeros will be the reverse of those shown.

(3.2) Optimization of Gain/Rise-Time Quotients

Using these results, it is now possible to obtain expressions for gain/rise-time quotients, based on Elmore's rise-time definition, in terms of pole and zero positions, and hence an optimum pole-zero pattern can be obtained which yields the best possible gain/rise-time under the given physical restrictions. These restrictions are as follows:

(a) The poles and zeros must be either real or occur in conjugate complex pairs and must all be confined to the left half of the complex frequency plane.
(b) The real part of the system function must be greater than, or equal to, zero for all real frequencies.
(c) The poles and zeros must lie in regions corresponding to monotonic response.[6,7,8]
(d) The system function must tend to $1/pC$ (where C is the parasitic shunt capacitance), as the complex frequency tends to infinity.

Since real poles will appear in a symmetrical fashion both in the gain/rise-time quotient and in the mathematical expressions corresponding to the various constraints, it follows that the

optimum gain/rise-time quotient will occur when all the real poles are coincident. Similarly, the optimum gain/rise-time quotient demands that all real zeros should be coincident, all pairs of conjugate complex poles coincident, and all pairs of conjugate complex zeros coincident.

Applying this method it can be demonstrated that the critically-damped shunt-peaked circuit[9] gives the optimum gain/rise-time quotient consistent with monotonic response for a circuit complexity of one zero and two poles. The maximum gain/rise-time quotient obtainable with circuits containing three real poles is discussed in Section 3.3.

(3.3) Maximum Value of Figure of Merit obtainable with Three Real Poles, Two Real Zeros

Let p_1' indicate the position of the triple pole, and p_1 that of the double zero. Then the system function, $Z(p)$, is given by

$$Z(p) = G\frac{(1 - p/p_1)^2}{(1 - p/p_1')^3} \quad . \quad . \quad . \quad (14)$$

The constraint imposed by the parasitic shunt capacitance is

$$GC = -p_1^2/p_1'^3 \quad . \quad . \quad . \quad . \quad (15)$$

But

$$\frac{T_R^2}{2\pi} = \frac{3}{p_1'^2} - \frac{2}{p_1^2}$$

Hence

$$\text{Gain/rise-time} = \frac{1}{C\sqrt{(2\pi)}} \frac{1}{\sqrt{\left[3\left(\frac{p_1'}{p_1}\right)^4 - 2\left(\frac{p_1'}{p_1}\right)^6\right]}} \quad . \quad (16)$$

However, the gain/rise-time quotient, for the simplest possible 2-terminal network consisting of parallel RC, is easily shown to be $1/[C\sqrt{(2\pi)}]$, when Elmore's definition of rise-time is taken. Thus, if Φ is the figure of merit using Elmore's rise-time definition, then

$$\Phi^2 = \frac{1}{3\left(\frac{p_1'}{p_1}\right)^4 - 2\left(\frac{p_1'}{p_1}\right)^6} \quad . \quad . \quad . \quad (17)$$

or

$$\Phi^2 = \frac{1}{3x^2 - 2x^3}$$

where

$$x = (p_1'/p_1)^2$$

and, for monotonic response, $x < 1$.

Now

$$Z(j\omega) = G\frac{(1 - j\omega/p_1)^2}{(1 - j\omega/p_1')^3} \quad . \quad . \quad . \quad (18)$$

and hence

$$\text{Real part of } Z(j\omega) = \frac{G}{\left[1 + \frac{\omega^2}{p_1'^2}\right]^2}\left[1 - \omega^2\left(\frac{1}{p_1^2} + \frac{3}{p_1'^2} - \frac{6}{p_1 p_1'}\right)\right.$$
$$\left. + \omega^4\left(\frac{3}{p_1^2 p_1'^2} - \frac{2}{p_1 p_1'^3}\right)\right] \quad . \quad . \quad (19)$$

and so the passive restriction demands that

$$\frac{3}{p_1^2 p_1'^2} \geqslant \frac{2}{p_1 p_1'^3}$$

i.e.

$$p_1'/p_1 \geqslant 2/3$$

or

$$x \geqslant 4/9$$

Further, since $\frac{1}{p_1^2} + \frac{3}{p_1'^2} - \frac{6}{p_1 p_1'}$ is negative over the range $1 > p_1'/p_1 \geqslant 2/3$, it follows that this restriction is also sufficient for passive realizability. The maximum value of Φ has therefore to be determined over the range $1 > x \geqslant 4/9$.

It is easily seen that $x = 4/9$ yields the maximum value of Φ, given by

$$\Phi = \frac{27}{4\sqrt{19}} = 1\cdot55$$

which represents a very small improvement over the corresponding figure of $1\cdot51$ for the shunt-peaked circuit.

(3.4) An Upper Limit to the Figure of Merit obtainable with n Real Poles and $n - 1$ Real Zeros as $n \to \infty$

In the case of n real poles and $n - 1$ real zeros,

$$Z(p) = G\frac{(1 - p/p_1)^{n-1}}{(1 - p/p_1')^n} \quad . \quad . \quad . \quad (20)$$

Imposing the constraint of the parasitic shunt capacitance, the figure of merit is finally obtained as

$$\Phi^2 = \frac{1}{nx^{n-1} - (n - 1)x^n} \quad . \quad . \quad . \quad (21)$$

where $x = (p_1'/p_1)^2$ as in Section 3.3, and $x < 1$ yields monotonic response.

A limit to the value of x imposed by the passive condition can be obtained by examining the condition for the sign of the highest power of ω, namely $2n - 2$, in the numerator of the expression for the real part of $Z(j\omega)$, to be positive. This will be found to yield $x \geqslant [(n - 1)/n]^2$, and, inserting $x = [(n - 1)/n]^2$ in the expression for Φ^2, we have

$$\Phi^2 = \frac{1}{n\left(\frac{n - 1}{n}\right)^{2n-2} - (n - 1)\left(\frac{n - 1}{n}\right)^{2n}}$$

$$= \frac{1}{3(1 - 1/n)^{2n-1} + \frac{1}{n^2}(1 - 1/n)^{2n-2}}$$

$$= \frac{\varepsilon^2}{3 + \frac{1}{2n^2} + \frac{1}{2n^3} + \frac{7}{8n^4} + \text{etc.}} \quad . \quad . \quad . \quad (22)$$

so that the upper limit to the value of the figure of merit in this case is

$$\Phi = \frac{\varepsilon}{\sqrt{3}} = 1\cdot569$$

This represents only a very slight improvement over the previous case. Better values can be obtained by taking complex poles and zeros. It will be found that with three poles and two zeros the maximum figure of merit is obtained by taking both a pair of conjugate complex poles and a pair of conjugate complex zeros. For this reason, only this case will be considered here.

(3.5) Maximum Value of Figure of Merit obtainable with a Pair of Conjugate Complex Zeros and Three Poles, Two of which are Conjugate Complex

Let the poles and zeros be numbered as in Fig. 4.
The system function is given by

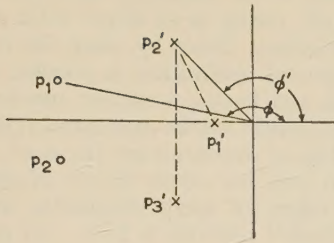$$Z(p) = G\frac{(1 - p/p_1)(1 - p/p_2)}{(1 - p/p_1')(1 - p/p_2')(1 - p/p_3')} \quad . \quad . \quad (23)$$

Fig. 4.—Numbering scheme for poles and zeros in Section 3.5.

The constraint imposed by the parasitic shunt capacitance is

$$GC = \frac{|p_1|^2}{-p_1' \times |p_2'|^2} \quad . \quad . \quad . \quad . \quad (24)$$

and the Elmore rise-time is given by

$$\frac{T_R^2}{2\pi} = \frac{1}{p_1'^2} + \frac{2\cos 2\phi'}{|p_2'|^2} - \frac{2\cos 2\phi}{|p_1|^2} \quad . \quad . \quad . \quad (25)$$

Thus the Elmore figure of merit, Φ, is given by

$$\Phi^2 = \frac{1}{x^4 + 2x^2 y^2 \cos 2\phi' - 2x^4 y^2 \cos 2\phi} \quad . \quad (26)$$

where, in this Section,

$$x = \frac{|p_2'|}{|p_1|}, \quad x > 0$$

$$y = \frac{p_1'}{|p_1|}, \quad y < 0, \quad x^2 > y^2$$

and the problem is to maximize Φ within the restrictions of passive realizability and monotonic response, or to minimize F where

$$F = 1/\Phi^2 = x^4 + 2x^2 y^2 \cos 2\phi' - 2x^4 y^2 \cos 2\phi \quad . \quad (27)$$

For the passive restrictions, proceeding as before by rationalizing $Z(j\omega)$,

$$\text{Real part of } Z(j\omega) = \frac{1 - a\omega^2 + b\omega^4}{(\text{a quantity always positive for real } \omega)} \quad . \quad (28)$$

where $a = \dfrac{1}{|p_1|^2} + \dfrac{1}{|p_2'|^2} + \dfrac{2\cos\phi'}{|p_2'||p_1'|} - \dfrac{2\cos\phi}{|p_1||p_1'|} - \dfrac{4\cos\phi\cos\phi'}{|p_1||p_2'|}$

and $b = \dfrac{1}{|p_1|^2|p_2'|^2} + \dfrac{2\cos\phi'}{|p_1|^2|p_2'||p_1'|} - \dfrac{2\cos\phi}{|p_2'|^2|p_1||p_1'|}$

Thus for the real part of $Z(j\omega)$ to be non-negative for all real ω it is necessary that $b > 0$ and either $a < 0$ or, if not, then $a^2 \leqslant 4b$. In terms of the variables x and y this last inequality becomes

$$\left(x + \frac{1}{x} + \frac{2\cos\phi'}{y} - \frac{2x\cos\phi}{y} - 4\cos\phi\cos\phi'\right)^2$$
$$\leqslant 4\left(1 + \frac{2x\cos\phi'}{y} - \frac{2\cos\phi}{y}\right) \quad . \quad (29)$$

As pointed out by Mulligan,[6] a necessary condition for monotonic response is that, in the present notation, p_1' should not lie to the left of the line joining p_2' and p_3',

i.e. $$y^2/x^2 \leqslant \cos^2\phi'$$

Inserting the limiting condition $y^2/x^2 = \cos^2\phi'$, expressions (27) and (29) become

$$F = x^4 + 4y^4 - 2x^2 y^2 - 2x^4 y^2 \cos 2\phi \quad . \quad . \quad (30)$$

and $$\left[x + \frac{3}{x} - 2\cos\phi\left(\frac{x}{y} + \frac{2y}{x}\right)\right]^2 \leqslant 4\left(3 - \frac{2\cos\phi}{y}\right) \quad . \quad (31)$$

The usual method of determining constrained maxima and minima, i.e. Lagrange's method of undetermined multipliers, results in such complicated algebra when applied to these two expressions that recourse was had instead to a purely numerical method. Various values were successively allocated to $\cos 2\phi$, and, for each of these values, x and y were varied over their allowable ranges, the combination of values yielding the minimum F being noted. The maximum figure of merit Φ ($= 1/\sqrt{F}$) obtained was $1\cdot98$, given by $\cos 2\phi = 0\cdot2$, $y = -0\cdot535$, $x = 0\cdot688$. However, the optimum value of Φ is very insensitive to changes in $\cos 2\phi$ around this value, e.g. maximum Φ for $\cos 2\phi = 0\cdot1$ was $1\cdot975$, given by $y = -0\cdot525$, $x = 0\cdot692$, and maximum Φ for $\cos 2\phi = 0\cdot5$ was $1\cdot90$, given by $y = -0\cdot578$, $x = 0\cdot707$. This would give a certain degree of freedom in the actual physical realization of the network.

In terms of the quantities x, y and $\cos\phi$ the system function can be written:

$$Z(p) = G\frac{1 + 2x^2 y\cos\phi pGC + x^4 y^2 p^2 G^2 C^2}{1 + (x^2 + 2y^2)pGC + 3x^2 y^2 p^2 G^2 C^2 + x^4 y^2 p^3 G^3 C^3} \cdot (32)$$

Taking out the parasitic shunt capacitance C,

$$Y(p) = 1/Z(p) = pC + Y_1(p) \quad . \quad . \quad . \quad (33)$$

where

$$Y_1(p) = \frac{1 + (x^2 + 2y^2 - 1)pGC + (3x^2 y^2 - 2x^2 y\cos\phi)p^2 G^2 C^2}{G + 2x^2 y\cos\phi pG^2 C + x^4 y^2 p^2 G^3 C^2}$$
$$. \quad . \quad . \quad (34)$$

The driving-point admittance $Y_1(p)$ is a minimum conductance function[10] (provided that x, y and $\cos\phi$ represent a limiting combination just satisfying the passive restriction), and the corresponding physical network has now to be determined. One such realization, obtained by Brune's method,[10,11] is shown in Fig. 5.
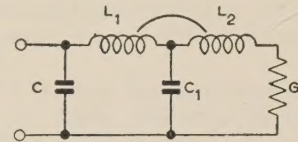

Fig. 5.—Network realized by Brune's method.

$$L_1 = \frac{x^4 y^2}{x^2 + 2y^2 - 1}G^2 C$$

$$L_2 = \frac{3x^2 y^2 - 2x^2 y\cos\phi}{x^2 + 2y^2 - 1}G^2 C$$

$$C_1 = (x^2 + 2y^2 - 1)C$$

100% coupling between L_1 and L_2, connected in series opposing.

Inserting the values $\cos 2\phi = 0\cdot2$, $y = -0\cdot535$, $x = 0\cdot688\,19$, in eqn. (34),

$$Y_1(p) = \frac{1 + 0\cdot046\,06pGC + 0\cdot014\,14p^2 G^2 C^2}{G + 0\cdot392\,54pG^2 C + 0\cdot064\,20p^2 G^3 C^2} \cdot (35)$$

and the corresponding numerical values in the Brune realization are $L_1 = 1\cdot394G^2 C$, $L_2 = 0\cdot307G^2 C$ and $C_1 = 0\cdot0461C$. These values would not perhaps be suitable for some practical applications, in which case the parameters x, y and $\cos\phi$ could be varied somewhat from the optimum, producing a more suitable practical realization with very little reduction in the figure of merit. For example, the value of the network capacitance in terms of the parasitic shunt capacitance, $x^2 + 2y^2 - 1$, could be raised considerably by small increases in x^2 and y^2.

Brune's method of realizing a driving-point impedance always yields perfectly coupled coils, and it has usually been considered that this renders the realization of theoretical interest only. However, with the magnetic core materials now available, very high coefficients of coupling, approaching unity, can be obtained, which may now result in a much greater practical use of Brune's method than has hitherto been the case. Three other methods[12,13,14] of realizing a driving-point impedance have recently been published.

On taking the inverse transform of the expression for $Z(p)$, the normalized transient response to a unit step of current, $f_n(t)$, is finally obtained as

$$f_n(t) = 1 - \exp\left(-\frac{t}{x^2 GC}\right)\left\{\frac{x^2(1 - 2y\cos\phi + y^2)}{x^2 - y^2}\right.$$

$$- \frac{y(-y - x^2 y + 2x^2 \cos\phi)}{y^2 - x^2}\cos\left[\frac{\sqrt{(x^2 - y^2)}}{x^2 y}\frac{t}{GC}\right]$$

$$\left. + \frac{y(x^2 - 1)\sqrt{(x^2 - y^2)}}{y^2 - x^2}\sin\left[\frac{\sqrt{(x^2 - y^2)}}{x^2 y}\frac{t}{GC}\right]\right\} \quad . \quad (36)$$

Inserting $\cos 2\phi = 0\cdot 2$, $y = -0\cdot 535$ and $x = 0\cdot 688$, the transient response becomes

$$f_n(t) = 1 - \exp\left(-2\cdot 113\frac{t}{GC}\right)\left(1\cdot 157 - 0\cdot 157\cos 1\cdot 71\frac{t}{GC}\right.$$

$$\left. + 0\cdot 651\sin 1\cdot 71\frac{t}{GC}\right) \quad . \quad (37)$$
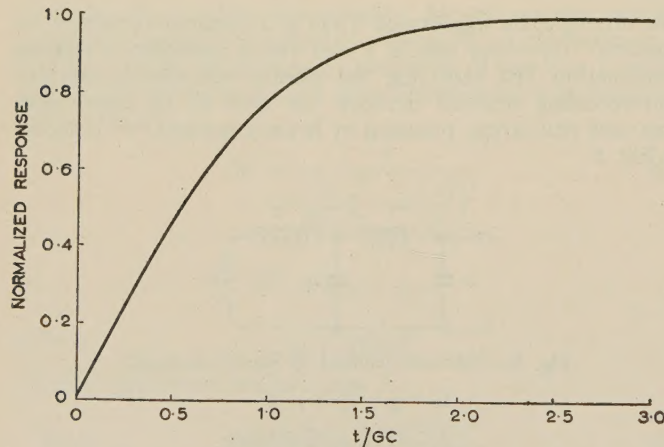
which is shown plotted in Fig. 6.

Fig. 6.—Transient response of optimum network.

The 10%–90% rise-time is $1\cdot 25GC$, which corresponds to a figure of merit, using the 10%–90% rise-time definition, of $1\cdot 76$.

In the process of optimizing the Elmore figure of merit, Φ, decreasing y to somewhat below its maximum allowable value of $(2\cos\phi)/3$ was found to increase considerably the value of Φ. For example, for $\cos 2\phi = 0\cdot 2$, the maximum value of y is $-0\cdot 516$, which, together with the appropriate x^2 value of $0\cdot 7$, yields $\Phi = 1\cdot 69$, whereas $y = -0\cdot 535$ and $x = 0\cdot 688$ yield $\Phi = 1\cdot 98$. However, the 10%–90% figures of merit for these two values of y are practically identical, being $1\cdot 74$ and $1\cdot 76$, respectively. The curve for the transient response corresponding to the maximum allowable value of y does not differ sufficiently from the optimum curve, shown in Fig. 6, to enable it to be reproduced in the same diagram. The two curves are practically identical up to about 90% of their final value, and thereafter the curve corresponding to the maximum allowable value of y has a rather longer 'tail', owing to its larger value of x and hence smaller decay coefficient. Since by using the maximum value of y a simpler network realization is possible, without using mutual inductance and having a smaller number of elements, only this series of networks will be considered in the next Section, in which small values of overshoot are discussed.

For comparison with the above figures of merit, the maximum 10%–90% figure of merit obtainable with monotonic response with a Doba[15] network is $1\cdot 68$. As originally given, this network was said to have a 10%–90% figure of merit of $1\cdot 77$ with 1% overshoot. Since a total of four elements are used to synthesize a function having three poles and two zeros, two constraints are present, which can be taken to imply that the zeros are determined if the poles and gain level are given. On inserting the limiting condition of the poles for monotonic response, the maximum 10%–90% figure of merit obtainable is $1\cdot 68$.

(4) TWO-TERMINAL NETWORKS HAVING MINIMUM RISE-TIME CONSISTENT WITH A SMALL DEGREE OF OVERSHOOT

Only networks having a pair of conjugate complex zeros and three poles (two of which are conjugate complex) are considered in this Section. The results deduced for monotonic response in Section 3.5 are here modified to allow for a small degree of overshoot. Using the notation of Section 3.5, but writing k for $(x\cos\phi')/y$, the system function $Z(p)$, and the normalized transient response to a unit step of current, $f_n(t)$, can be obtained as

$$Z(p) = G\frac{(1 + 2x^2 y\cos\phi pGC + x^4 y^2 p^2 G^2 C^2)}{(1 + x^2 pGC)(1 + 2ky^2 pGC + x^2 y^2 p^2 G^2 C^2)} \quad . \quad (38)$$

and

$$f_n(t) = 1 - \exp\left(-\frac{k}{x^2}\frac{t}{GC}\right)\left\{\frac{x^2(1 - 2y\cos\phi + y^2)}{x^2 - y^2(2k - 1)}\right.$$

$$\times \exp\left[-\frac{(1 - k)}{x^2}\frac{t}{GC}\right]$$

$$- \frac{(2k - 1)y^2 + x^2 y^2 - 2x^2 y\cos\phi}{x^2 - y^2(2k - 1)}\cos\frac{1}{x^2}\frac{t}{GC}\sqrt{(x^2/y^2 - k^2)}$$

$$+ \frac{x^2(1 - x^2) - 2x^2 y(1 - k)\cos\phi + ky^2(x^2 + 1) - 2k^2 y^2}{\sqrt{(x^2/y^2 - k^2)}[x^2 - y^2(2k - 1)]}$$

$$\left. \times \sin\frac{1}{x^2}\frac{t}{GC}\sqrt{(x^2/y^2 - k^2)}\right\} \quad . \quad (39)$$

In Section 3.5 k was taken to be equal to unity, the limiting position for monotonic response. For $k < 1$, the real pole lies to the left of the line joining the pair of conjugate complex poles and the response must be non-monotonic. However, as k is reduced below unity, the rise-time can be considerably reduced without at first occasioning a large amount of overshoot. If the maximum value of y [$= 2\cos\phi/(1 + 2k)$], is selected so as to yield the simplest series of networks, and $\cos 2\phi$ is taken as $0\cdot 2$, then the improvement in 10%–90% figure of merit as k is reduced from $1\cdot 0$ to $0\cdot 7$ is as shown in Fig. 7, together with the corresponding increase in overshoot; the curve showing overshoot has a discontinuity at about $k = 0\cdot 845$, and the appropriate values of x are obtained from the equation

$$x^2 = \frac{2k + 1 - \dfrac{8k\cos^2\phi}{2k + 1}}{2k} \quad . \quad . \quad . \quad (40)$$
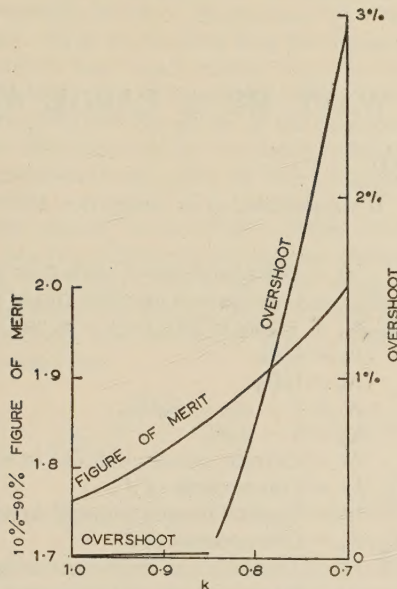
Fig. 7.—Improvement in figure of merit with increase in overshoot.
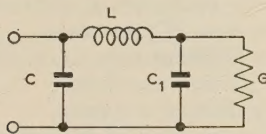

Fig. 8.—Realization of series of networks having maximum value of y.

$$L = \frac{x^4 y^2}{x^2 + 2ky^2 - 1} G^2 C$$
$$= (2k + 1)x^2 y^2 G^2 C$$
$$C_1 = (x^2 + 2ky^2 - 1)C$$
$$= \frac{x^2}{2k + 1} C$$

Since

$$\frac{1}{Z(p)} = pC + \cfrac{1}{pG^2 C \dfrac{x^4 y^2}{x^2 + 2ky^2 - 1} + \cfrac{G}{pGC(x^2 + 2ky^2 - 1) + 1}}$$

$$. \quad . \quad . \quad . \quad (41)$$

the series of networks can be realized in the form shown in Fig. 8. The physical realization of these networks can be varied some-

what without appreciably affecting the rise-time for a given value of overshoot by either

(a) altering the value of $\cos \phi$, which would alter the induc-
tance and capacitance values,

or

(b) adding resistance in series with the inductance, which
corresponds to using a value of y below the maximum.

(5) REFERENCES

(1) MULLER, F. A.: 'High-Frequency Compensation of *RC* Amplifiers', *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 1271.

(2) VAN DER POL, B., and BREMMER, H.: 'Operational Calculus based on the Two-sided Laplace Integral' (Cambridge University Press, 1950).

(3) *Ibid.*, p. 171.

(4) MATHERS, G. W. C.: 'Technical Report No. 38' (Electronics Research Laboratory, Stanford University, California, November, 1951).

(5) ELMORE, W. C.: 'The Transient Response of Damped Linear Networks with Particular Regard to Wide-Band Amplifiers', *Journal of Applied Physics*, 1948, **19**, p. 55.

(6) MULLIGAN, J. H.: 'The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems', *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 516.

(7) MACDIARMID, I. F.: 'Transient Response', *Wireless Engineer*, 1951, **28**, p. 330.

(8) CUTTERIDGE, O. P. D.: 'Monotonic Transient Response', *Proceedings I.E.E.*, Monograph No. 74 R, September, 1953 (**101**, Part IV, p. 46).

(9) VALLEY, G. E., and WALLMAN, H.: 'Vacuum Tube Amplifiers', M.I.T. Radiation Laboratory Series, **18** (McGraw-Hill, New York, 1948), p. 73.

(10) BODE, H. W.: 'Network Analysis and Feedback Amplifier Design' (D. Van Nostrand, New York, 1945).

(11) BRUNE, O.: 'Synthesis of a Finite Two-Terminal Network whose Driving-Point Impedance is a Prescribed Function of Frequency', *Journal of Mathematics and Physics*, 1931, **10**, p. 191.

(12) MIYATA, F.: 'A New System of Two-Terminal Synthesis', *Institute of Radio Engineers Transactions on Circuit Theory*, 1955, CT–2, p. 297.

(13) KUH, E. S.: 'Special Synthesis Techniques for Driving-Point Impedance Functions', *ibid.*, p. 302.

(14) SESHU, S.: 'Topological Considerations in the Design of Driving-Point Functions', *ibid.*, p. 356.

(15) VALLEY, G. E., and WALLMAN, H.: *op. cit.*, pp. 74–75.

621.396.11 : 538.566

THE DIFFRACTION OF AN ELECTROMAGNETIC WAVE BY A LARGE APERTURE

By R. F. MILLAR, M.A.

SUMMARY

With the aid of methods previously developed,[1,2] an approximate expression is obtained for the difference from unity of the transmission coefficient for plane waves at normal incidence of a large aperture of rather general form in a plane perfectly-conducting screen. The coefficient is expressed in the form of one or more line integrals taken around all (or part) of the aperture rim. These may be evaluated analytically in certain cases, e.g. a circular or elliptical aperture. In the former case the result agrees with that derived elsewhere;[2] the ellipse is considered in some detail in the present paper, and the transmission coefficient is evaluated for a range of eccentricities. The results appear to be in qualitative agreement with the limited experimental data available.

LIST OF SYMBOLS

(x, y, z), (X, Y, z), (X_0, Y_0, z) = Systems of rectangular Cartesian co-ordinates.

λ = Wavelength of incident radiation.

ω = Angular frequency of incident radiation.

c = Speed of light in free space.

$k = \omega/c$.

ϵ_0 = Permittivity of free space.

μ_0 = Permeability of free space.

$Z_0 = \sqrt{(\mu_0/\epsilon_0)}$.

$E^i(z)$ = Normally incident electric field of unit intensity.

E_x^d, E_y^d = Diffracted components of electric field.

E^{int} = Interaction component of electric field in aperture.

E_x^{int}, E_y^{int} = Components of E^{int}.

E^p = Plane-wave approximation to E^{int}.

E_x^p, E_y^p = Components of E^p.

E^0 = Electric field incident obliquely on half-plane.

E_0, E_1 = Constants of proportionality.

H_x = x-component of magnetic field.

P = Point in aperture.

Q, Q', Q_0, Q_0' = Points on rim of aperture.

ρ = Distance between P and Q.

ρ_0' = Distance between P and Q_0'.

ρ_0 = Distance between Q_0 and Q_0'.

ρ_1 = Distance between Q_0 and Q_0' when Q_0 and Q_0' lie at ends of principal axis.

s = Arc length measured along rim of aperture.

s_0, s_0' = Values of s at Q_0, Q_0', respectively.

ds = Element of rim of aperture at Q.

γ = Angle between positive x-axis and PQ.

θ = Angle between $E^i(0)$ and ds.

θ_0, θ_0' = Values of θ at Q_0, Q_0', respectively.

δ = Angle between axes of X_0 and X.

$m = \sin \delta$.

$l = -\cos \delta$.

ϕ = Polar angle of Q, with pole at P.

$K(s)$ = Curvature of aperture rim as a function of s.

K_0, K_0' = Values of $K(s)$ for $s = s_0$ and s_0', respectively.

$r_0 = 1/K_0$.

$r_0' = 1/K_0'$.

$\chi' = 1 - (\rho_0 + X_0)K_0'$.

$\chi_0' = 1 - \rho_0 K_0'$.

J^0 = Current density induced in screen by E^0.

J_X^0, J_Y^0 = Components of J^0.

J = Current density induced in screen by E^p.

J_X, J_Y, J_y = Components of J.

t = Transmission coefficient of aperture.

t' = Interaction contribution to t.

t_c' = Interaction contribution to transmission coefficient of circular aperture.

t_e' = Interaction contribution to transmission coefficient of elliptical aperture.

A = Area of aperture.

R = Distance from (x, y, z) to point of integration on screen.

S = Surface of screen.

$\Pi_y(x, y, z)$ = y-component of Hertz vector with source J.

$\Psi = \rho_1 K_0 K_0' - K_0 - K_0'$.

$\Phi_0 = \dfrac{d^2\rho_0(s)}{ds^2}\bigg]_{s=s_0}$.

$\Phi_0' = \dfrac{d^2\rho_0(s)}{ds^2}\bigg]_{s=s_0'}$.

$2a$ = Length of major axis of ellipse.

$2b$ = Length of minor axis of ellipse.

$e = [1 - (b/a)^2]^{1/2}$ = Eccentricity of ellipse.

ϕ_1 = Angle between x-axis and major axis of ellipse.

$\Gamma(z)$ = Gamma function of argument z.

(r, ϕ) = Polar co-ordinates of point on ellipse with pole at centre.

$J_n(x)$ = Bessel function of the first kind of order n and argument x.

Rationalized M.K.S. units are employed throughout.

(1) INTRODUCTION

In a recent paper[1] an approximate method was described for examining the diffraction of a normally incident plane wave by an aperture in a plane perfectly-conducting screen of negligible thickness. Since this underlies the following development, the basic principles involved are briefly restated.

Consider the solution to the 2-dimensional problem of the diffraction of an incident plane wave by a perfectly conducting half-plane.[3] It appears that the diffracted* far-field may be derived, except in certain critical regions, from fictitious electric and magnetic currents flowing on the edge of the half-plane. Associated with each current is a polar diagram, so the magnitude of the current depends upon the point from which it is observed.

* The diffracted field is the difference between the total field and that predicted by geometrical optics.

This concept is readily extended to the consideration of the diffraction of normally incident plane waves by a large aperture in a plane screen—large in the sense that the radius of curvature of the rim is always very much greater than the wavelength of the incident radiation. Then the aperture rim may be considered to be locally straight, and the screen in the neighbourhood of a point on the rim approximated by a correctly oriented half-plane. The far-field components are given as line integrals around the aperture rim.

If the incident electric vector $E^i(z) = E^i(0)\varepsilon^{-jkz}$ is polarized parallel to the positive y-axis of a rectangular Cartesian co-ordinate system (x, y, z), and $E^i(0) = (0, 1, 0)$, while $z = 0$ is the plane of the aperture and screen, the diffracted component of the tangential electric field at a point $P(x, y, 0)$ in the aperture takes the following form:

$$
\left.
\begin{aligned}
E_x^d(x, y, 0) &= \frac{1}{\sqrt{2\pi}} \oint \frac{\varepsilon^{-jk\rho}}{\rho} \cos(\gamma - \theta) \cos\theta \sin\gamma \, ds \\[2mm]
E_y^d(x, y, 0) &= -\frac{1}{\sqrt{2\pi}} \oint \frac{\varepsilon^{-jk\rho}}{\rho} \cos(\gamma - \theta) \cos\theta \cos\gamma \, ds
\end{aligned}
\right\} \quad (1)
$$

where the various symbols are illustrated in Fig. 1. (The common time-dependence factor $\varepsilon^{j\omega t}$ is suppressed throughout.)
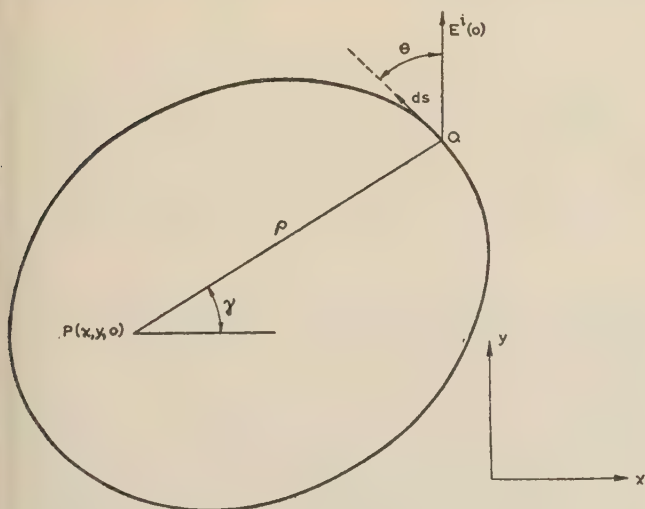

Fig. 1.—Aperture in plane screen, with associated symbols.

Implicit in the derivation of eqns. (1) is the assumption that the radiation from each current element on the rim to the point P is not intercepted by any portion of the screen, or, equivalently, that the curve defining the aperture rim is everywhere concave inwards (i.e. it has positive curvature). Considerable complication arises if this restriction is relaxed. Positive curvature of the rim is assumed in the following analysis, although the results obtained may be also valid for some apertures of more general form.

The method developed in the present paper is essentially an extension of that developed for, and applied to, the particular problem of diffraction by a circular aperture.[2] The integrals in eqns. (1) which define the aperture field are evaluated asymptotically to the first order in $1/k^{1/2}$ by the principle of stationary phase. The field at a point P near the aperture rim is then seen to consist of a component attributable to currents in the screen near P, together with contributions from across the aperture. These latter will be termed the 'interaction components' of the aperture field. By approximating the screen

near P by a correctly oriented half-plane, the first-order modification to the far field in the direction of propagation of the incident wave due to interaction is found in the form of line integrals taken around all or part of the aperture rim. The cross-section theorem[4] then gives the transmission coefficient, t, of the aperture in terms of these integrals.

The explicit determination of t would, in general, prove extremely difficult, but in certain cases the integrals may be evaluated asymptotically, again by the principle of stationary phase. Perturbation methods are applicable when the aperture is nearly circular in shape. Application of the former method yields a general asymptotic formula for the interaction component of the transmission coefficient in terms of certain 'principal axes' of the aperture. These will be defined later as they arise.

In particular, such an expression is found for the transmission coefficient of an elliptical aperture for which the eccentricity, e, lies not too near zero or unity. When e is very small the perturbation method is applied, and when e tends to zero it yields the expression for a circular aperture.[2] Finally, a comparison is made with measurements of the transmission coefficients of elliptical apertures of eccentricities $\sqrt{3}/2$ and $(2\sqrt{2})/3$ as functions of the lengths of the major axes.[5]

(2) THE INTERACTION COMPONENT OF THE APERTURE FIELD

Consider a point P in the aperture, close to the rim, as illustrated in Fig. 2. Q is a variable point on the rim which is
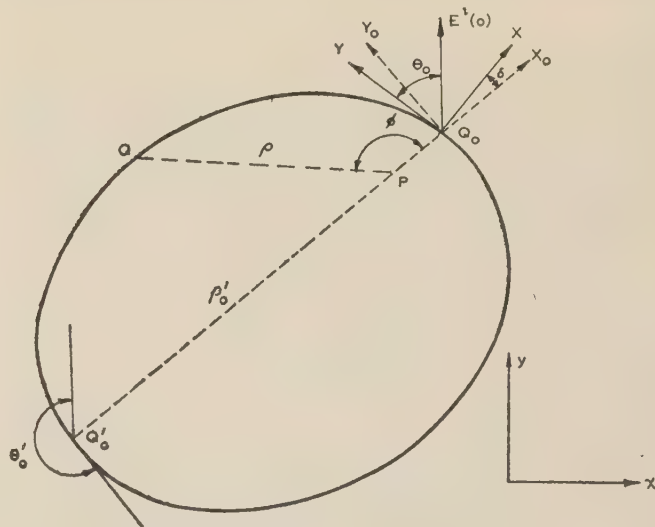

Fig. 2.—Aperture in plane screen, showing stationary point, Q_0', of ρ, and co-ordinate systems with origins at Q_0.

described parametrically by the arc length s, so that $Q = Q(s)$. The distance PQ, or ρ, is then, for fixed P, a function of s.

Suppose that $d\rho/ds = 0$ for $s = s_0'$, where $Q(s_0') \equiv Q_0'$, and P, if allowed to approach the rim, does not tend to coincide with Q_0'. PQ_0' is then perpendicular to the tangent at Q_0'. (In general there will be several such points, each of which must be considered separately.) The line $Q_0'P$ meets the rim again in $Q_0 \equiv Q(s_0)$. Q_0 is taken as the origin of two rectangular Cartesian co-ordinate systems:

(a) An (X, Y, z) system in which X is directed normally to the rim at Q_0, away from the aperture, and Y is in the direction of the positive tangent at Q_0;

(b) An (X_0, Y_0, z) system in which the positive X_0-axis is the extension of PQ_0. The direction of the positive Y_0-axis is such that (X_0, Y_0, z) forms an orthogonal, right-handed system. The

point P is thus described by the co-ordinates $(X_0, 0, 0)$ in the (X_0, Y_0, z) system.

$Q(s)$ may also be specified by the plane polar co-ordinates (ρ, ϕ), where the pole is at P and $\phi = 0$ corresponds to PQ_0. Then

$$\left.\begin{array}{l}\rho = \rho(\phi) \\ \phi = \phi(s)\end{array}\right\} \quad . \quad . \quad . \quad . \quad (2)$$

$$\left.\begin{array}{l}\dfrac{d\rho}{ds} = \dfrac{d\rho}{d\phi}\dfrac{d\phi}{ds} \\[2mm] \dfrac{d^2\rho}{ds^2} = \dfrac{d^2\rho}{d\phi^2}\left(\dfrac{d\phi}{ds}\right)^2 + \dfrac{d\rho}{d\phi}\dfrac{d^2\phi}{ds^2}\end{array}\right\} \quad . \quad . \quad . \quad (3)$$

Since $\dfrac{d\rho}{d\phi} = 0$ for $\phi = \pi$, or $s = s_0'$, and

$$\left(\dfrac{ds}{d\phi}\right)^2 = \rho^2 + \left(\dfrac{d\rho}{d\phi}\right)^2 \quad . \quad . \quad . \quad . \quad (4)$$

the following is obtained:

$$\left.\dfrac{d^2\rho}{ds^2}\right]_{s=s_0'} = \dfrac{1}{(\rho_0')^2}\left.\dfrac{d^2\rho}{d\phi^2}\right]_{\phi=\pi} \quad . \quad . \quad . \quad (5)$$

where ρ_0' is the distance PQ_0'.

If the curvature of the rim at Q is denoted by $K(s)$,

$$K(s_0') \equiv K_0' = \dfrac{1}{\rho_0'} - \dfrac{1}{(\rho_0')^2}\left.\dfrac{d^2\rho}{d\phi^2}\right]_{\phi=\pi} \quad . \quad . \quad (6)$$

Hence

$$\left.\dfrac{d^2\rho}{ds^2}\right]_{s=s_0'} = \dfrac{1}{\rho_0'} - K_0' \quad . \quad . \quad . \quad . \quad (7)$$

The first term in the asymptotic expansion of eqns. (1) in inverse powers of $k^{1/2}$ may be obtained by the principle of stationary phase. The parts of the integrands which are slowly varying near $s = s_0'$ may be given their values at s_0', taken outside the sign of integration, and the limits extended from $-\infty$ to $+\infty$.

At $s = s_0'$, $\gamma = \theta = \theta_0'$, say. Near $s = s_0'$, ρ has the following form:

$$\rho \simeq \rho_0' + \rho''(s_0')\dfrac{(s - s_0')^2}{2} \quad . \quad . \quad . \quad (8)$$

where

$$\rho''(s_0') \equiv \left.\dfrac{d^2\rho}{ds^2}\right]_{s=s_0'} \quad . \quad . \quad . \quad . \quad (9)$$

Then, from eqns. (1), at P,

$$E_x^d \simeq \dfrac{1}{\sqrt{2\pi}}\sin\theta_0'\cos\theta_0'\dfrac{\varepsilon^{-jk\rho_0'}}{\rho_0'}\int_{-\infty}^{\infty}\exp\left[-jk\rho''(s_0')\dfrac{(s-s_0')^2}{2}\right]ds$$
$$+ R_x + 0\left(\dfrac{1}{k}\right) \quad (10)$$

i.e.

$$E_x^d \simeq \dfrac{\sin\theta_0'\cos\theta_0'}{\sqrt{\pi}k^{1/2}|1 - \rho_0'K_0'|^{1/2}}\dfrac{\varepsilon^{-jk\rho_0'}}{\sqrt{\rho_0'}}\exp\left[-j\pi/4\,\mathrm{sgn}\,(1-\rho_0'K_0')\right]$$
$$+ R_x + 0\left(\dfrac{1}{k}\right) \quad (11)$$

where

$$\mathrm{sgn}\,(v) = \left\{\begin{array}{ll}1, & v > 0 \\ -1, & v < 0\end{array}\right. \quad . \quad . \quad . \quad (12)$$

and R_x is the contribution from a stationary point near Q_0, which becomes singular as X_0 approaches zero.

It is to be noted that eqn. (10) is valid only if $1 - \rho_0'K_0' \neq 0$, or equivalently, $\rho''(s_0') \neq 0$. Special attention must be given in the neighbourhood of points P at which this is violated and focusing occurs. Such effects arise in the case of an aperture of elliptical form, when the eccentricity is near $1/\sqrt{2}$. Attention

will be paid to this later, but in the following analysis it will be assumed that $\rho''(s_0') \neq 0$ in the neighbourhood of points P under consideration.

Let

$$\rho_0 = \rho_0' - X_0 \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Then,

$$E_x^d \simeq \dfrac{\sin\theta_0'\cos\theta_0'}{\sqrt{\pi}k^{1/2}|X'|^{1/2}}\varepsilon^{-jk\rho_0}\exp\left(-j\pi/4\,\mathrm{sgn}\,X'\right)$$
$$\dfrac{\varepsilon^{-jkX_0}}{\sqrt{(\rho_0 + X_0)}} + R_x + 0\left(\dfrac{1}{k}\right) \quad (14)$$

and similarly,

$$E_y^d \simeq -\dfrac{\cos^2\theta_0'}{\sqrt{\pi}k^{1/2}|X'|^{1/2}}\varepsilon^{-jk\rho_0}\exp\left(-j\pi/4\,\mathrm{sgn}\,X'\right)$$
$$\dfrac{\varepsilon^{-jkX_0}}{\sqrt{(\rho_0 + X_0)}} + R_y + 0\left(\dfrac{1}{k}\right) \quad (15)$$

with

$$X' \equiv 1 - (\rho_0 + X_0)K_0' \quad . \quad . \quad . \quad (16)$$

The terms R_x, R_y, although singular at $X_0 = 0$, do not correctly describe the behaviour of the field due to currents in the screen near Q_0, since the integrals from which they are derived yield only a far-field approximation. On the other hand, if $k\rho_0 \gg 1$, the remaining terms in eqns. (13) and (14) should provide a valid approximation to the interaction electric field, E^{int} at P near the rim.

$$E_x^{int} \simeq$$
$$\left.\begin{array}{l}\dfrac{\sin\theta_0'\cos\theta_0'}{\sqrt{\pi}k^{1/2}|X'|^{1/2}}\varepsilon^{-jk\rho_0}\exp\left(-j\pi/4\,\mathrm{sgn}\,X'\right)\dfrac{\varepsilon^{-jkX_0}}{\sqrt{(\rho_0+X_0)}} + 0\left(\dfrac{1}{k}\right) \\[4mm] E_y^{int} \simeq \\[2mm] \dfrac{-\cos^2\theta_0'}{\sqrt{\pi}k^{1/2}|X'|^{1/2}}\varepsilon^{-jk\rho_0}\exp\left(-j\pi/4\,\mathrm{sgn}\,X'\right)\dfrac{\varepsilon^{-jkX_0}}{\sqrt{(\rho_0+X_0)}} + 0\left(\dfrac{1}{k}\right)\end{array}\right\}$$
$$. \quad . \quad . \quad (17)$$

The interaction electric field at P has no z-component, for the electric field of any current distribution in the screen has a vanishing normal component in the aperture.

The total current flowing in the screen may, to the present approximation, be split into two components: (*a*) the current flowing when interaction is neglected; and (*b*) the current induced in the screen by the interaction field [eqns. (17)]. Distribution (*a*) produces the geometrical-optic value of unity for the transmission coefficient, plus additional smaller terms depending on the aperture dimensions. For a large circular aperture it appears that the interaction correction is more important than these latter terms, and it will be assumed that the same is true for apertures of more general shape (see, for example, Reference 2 for a more complete discussion of this aspect of the problem).

The currents induced in the screen by the interaction field will modify the scattered field; in particular, the far field in the direction of propagation of the incident wave will be affected. This latter field is closely related to the transmission coefficient, t, of the aperture,[4] and knowledge of the current distribution (*b*) will determine t', the interaction contribution to t.

(3) THE INTERACTION CONTRIBUTION TO THE TRANSMISSION COEFFICIENT

In order to calculate the effect of the interaction field [eqns. (17)], it is once more assumed that the currents induced in the screen near Q_0 are approximately the same as those flowing in a similarly irradiated half-plane the edge of which lies along the tangent to the rim at Q_0. It will be found that these latter currents decay as the inverse square root of the distance from

the edge, rather than as the inverse distance, since the corresponding half-plane problem is only quasi 3-dimensional in character. However, this incorrect behaviour of the current density at large distances from the edge does not affect the far field in the direction of propagation of the incident wave to the first order in $1/k^{1/2}$.

It was stipulated in Section 2 that $\rho''(s_0')$ is not to vanish in the neighbourhood of points P under consideration. Since the current induced by the interaction field near the rim determines the above-mentioned first-order far-field, attention may be confined to this region. Thus it will be assumed that $\rho''(s_0') \neq 0$ for points P close to the edge of the aperture, or, equivalently, that X' does not change sign in the immediate neighbourhood of $X_0 = 0$.

The diffraction of the interaction field by the half-plane $z = 0$, $X \geqslant 0$, will now be considered (see Fig. 2). To a first approximation, the exciting field may be taken to be a plane wave, E^p, whose amplitude and phase at Q_0 are the same as the corresponding quantities in the interaction field.

$$E_x^p = \frac{\sin \theta_0' \cos \theta_0'}{\sqrt{\pi} k^{1/2} |X_0'|^{1/2}} \frac{\varepsilon^{-jk\rho_0}}{\sqrt{\rho_0}} \exp\left(-j\frac{\pi}{4}\operatorname{sgn} X_0'\right)\varepsilon^{-jkX_0} + 0\left(\frac{1}{k}\right)$$

$$E_y^p = -\frac{\cos^2 \theta_0'}{\sqrt{\pi} k^{1/2} |X_0'|^{1/2}} \frac{\varepsilon^{-jk\rho_0}}{\sqrt{\rho_0}} \exp\left(-j\frac{\pi}{4}\operatorname{sgn} X_0'\right)\varepsilon^{-jkX_0} + 0\left(\frac{1}{k}\right)$$

$$\cdots \cdots (18)$$

where
$$X_0' \equiv X'\Big]_{X_0=0} = 1 - \rho_0 K_0' \quad \cdots \cdots (19)$$

The terms designated as $0(1/k)$ will also contain the factor $(1/\sqrt{\rho_0}) \exp[-jk(\rho_0 + X_0)]$—a fact that will prove useful later.

The polarization of E^p is readily seen to be parallel to the Y_0-axis. With reference to the (X, Y, z) system, E^p then takes the following form:

$$E^p = E_0(-m, l, 0)\varepsilon^{jk(lX+mY)} \quad \cdots \quad (20)$$

where
$$E_0 = -\frac{\cos \theta_0'}{\sqrt{\pi} k^{1/2} |X_0'|^{1/2}} \frac{\varepsilon^{-jk\rho_0}}{\sqrt{\rho_0}} \exp\left(-j\frac{\pi}{4}\operatorname{sgn} X_0'\right) + 0\left(\frac{1}{k}\right) (21)$$

and
$$\left.\begin{array}{l} m = \sin \delta \\ l = -\cos \delta \\ \delta = \pi + \theta_0 - \theta_0' \end{array}\right\} \quad \cdots \cdots (22)$$

θ_0 is the value of θ corresponding to the point Q_0.

Consider then the field E^0 incident on $z = 0$, $X \geqslant 0$,

where
$$E^0 = (-m, l, 0)\varepsilon^{jk(lX+mY)} \quad \cdots \quad (23)$$

Copson's[6] solution to the quasi 3-dimensional problem of which this is a special case is provided in Section 9.1.

At a point $(X, 0, 0)$ the induced current density, J^0, equal in magnitude to the discontinuity in the tangential magnetic field across the screen, is readily found:

$$\left.\begin{array}{l} J_X^0(X, 0, 0) = 0 \\[2mm] Z_0 J_Y^0(X, 0, 0) = -\dfrac{2\sqrt{2}\varepsilon^{-j\pi/4}}{\sqrt{\pi} k^{1/2}|\cos \delta|^{1/2}}\dfrac{\varepsilon^{-jkX|\cos \delta|}}{\sqrt{X}} \end{array}\right\} \quad (24)$$

where
$$Z_0 = \sqrt{(\mu_0/\epsilon_0)} \quad \cdots \cdots (25)$$

Thus the current density, J, at $(X, 0, 0)$, induced by E^p, is given by the following expressions:

$$\left.\begin{array}{l} J_X(X, 0, 0) = 0 \\[2mm] Z_0 J_Y(X, 0, 0) = \dfrac{-2\sqrt{2}E_0}{\sqrt{\pi} k^{1/2}|\cos \delta|^{1/2}}\varepsilon^{-j\pi/4}\dfrac{\varepsilon^{-jkX|\cos \delta|}}{\sqrt{X}} + 0\left(\dfrac{1}{k^{3/2}}\right) \end{array}\right\}$$
$$\cdots \cdots (26)$$

Finally,

$$Z_0 J_y(X, 0, 0) = Z_0 J_Y(X, 0, 0)\cos \theta_0 = \frac{2\sqrt{2}\varepsilon^{-j\pi/4}}{\pi k}$$

$$\frac{\cos \theta_0 \cos \theta_0'}{|\cos \delta|^{1/2}|X_0'|^{1/2}}\frac{\varepsilon^{-jk\rho_0}}{\sqrt{\rho_0}}\exp\left(-j\frac{\pi}{4}\operatorname{sgn} X_0'\right)$$

$$\frac{\varepsilon^{-jkX|\cos \delta|}}{\sqrt{X}} + 0\left(\frac{1}{k^{3/2}}\right) \quad (27)$$

The contribution to the transmission coefficient due to interaction, t', is most readily found in terms of the magnetic far-field in the direction of propagation of the incident wave:

$$t' = -\lim_{z \to \infty} \frac{2\pi}{A}\mathscr{R}\left[\frac{Z_0 H_x(x, y, z)}{jk\varepsilon^{-jkz}/z}\right] \quad \cdots \quad (28)$$

where A is the area of the aperture.

If $\Pi_y(x, y, z)$ is the y-component of the Hertz vector whose source is J, then

$$\Pi_y(x, y, z) = \frac{1}{4\pi j\omega\epsilon_0}\iint_S J_y(X, 0, 0)\frac{\varepsilon^{-jkR}}{R}dS \quad (29)$$

where S represents the surface of the screen,

$$H_x(x, y, z) = -j\omega\epsilon_0\frac{\partial}{\partial z}\Pi_y(x, y, z) \quad \cdots \quad (30)$$

and R is the distance from (x, y, z) to the variable point of integration on the screen.

For $z > 0$,

$$\Pi_y(x, y, z) = \frac{1}{4\pi j\omega\epsilon_0}\frac{\varepsilon^{-jkz}}{z}\iint_S J_y(X, 0, 0)dS + 0\left(\frac{1}{z^2}\right) (31)$$

With the aid of eqns. (27)–(30) the following expression for t' is obtained:

$$t' = -\frac{\sqrt{2}}{\pi kA}\mathscr{R}\left\{\varepsilon^{-j\pi/4}\exp\left(-j\frac{\pi}{4}\operatorname{sgn} X_0'\right)\iint_S\left[\frac{\cos \theta_0 \cos \theta_0'}{|\cos \delta|^{1/2}|X_0'|^{1/2}}\right.\right.$$

$$\left.\left.\frac{\varepsilon^{-jk\rho_0}}{\sqrt{\rho_0}}\times\frac{\varepsilon^{-jkX|\cos \delta|}}{\sqrt{X}} + 0\left(\frac{1}{k^{1/2}}\right)\right]dS\right\} \quad (32)$$

If the radius of curvature of the aperture rim at Q_0 is $r_0 = 1/K_0$, then

$$dS = \left(1 + \frac{X}{r_0}\right)ds_0 dX \quad \cdots \cdots (33)$$

The integration with respect to X is of an elementary nature:

$$\int_0^\infty \frac{\varepsilon^{-jkX|\cos \delta|}}{\sqrt{X}}\left(1 + \frac{X}{r_0}\right)dX$$

$$= \frac{\sqrt{\pi}}{k^{1/2}|\cos \delta|^{1/2}}\varepsilon^{-j\pi/4}\left(1 - \frac{j}{2kr_0|\cos \delta|}\right) \quad (34)$$

where, as usual, it is assumed that k has a small negative imaginary part to ensure convergence. The integrand designated as $0(1/k^{\frac{1}{2}})$ in eqn. (32) will also contain $(1/\sqrt{X})\varepsilon^{-jkX|\cos \delta|}$ as a factor; on integration, the order therefore increases to $1/k$. Neglecting the term of order $1/k^{3/2}$ in eqn. (34), the term retained depends on the edge singularity of the current, but is unchanged if the current density is modified in such a way that it tends to zero as $1/X$ as $X \to \infty$, whilst retaining the present behaviour near $X = 0$.

Eqn. (31) has thus been reduced to a line integral along the aperture rim:

$$t' = -\frac{\sqrt{2}}{\sqrt{\pi} k^{3/2} A} \mathcal{R} \left\{ \exp\left[-j\pi/4 \left(\text{sgn} \, \chi_0' + 1\right) \right] \right.$$

$$\left. \times \int \left[\frac{\cos\theta_0 \cos\theta_0'}{|\cos\delta| \, |\chi_0'|^{1/2}} \frac{\varepsilon^{-jk\rho_0}}{\sqrt{\rho_0}} \varepsilon^{-j\pi/4} + 0\left(\frac{1}{k^{1/2}}\right) \right] ds_0 \right\} \quad (35)$$

In general, t' will consist of a sum of such integrals, each term being the contribution from one of the stationary points corresponding to Q_0. The range of integration over s_0 is in each case governed by the existence, or absence, of the related stationary point.

(4) REDUCTION OF THE LINE INTEGRAL

(4.1) The Circular Aperture

For a circular aperture of radius a there is but one term in the series for t', and the integration ranges around the entire aperture. $\theta_0' = \pi + \theta_0$, $\cos\delta = 1$, $\rho_0 = 2a$, and $\chi_0' = -1$. Thus t_c', the contribution due to interaction of the transmission coefficient for a circular aperture, is given by the following expression, also derived and discussed elsewhere:[2,7]

$$t_c' \simeq -\frac{1}{\sqrt{\pi}} \frac{\sin\left(2ka - \pi/4\right)}{(ka)^{3/2}} + 0\left[\frac{1}{(ka)^2}\right] \qquad . \quad (36)$$

(4.2) Apertures with Principal Axes

If ρ_0, considered as a function of s_0, has itself stationary values, then the line integral for t' may be evaluated asymptotically for $k \gg 1$, by the method of stationary phase. The line $Q_0 Q_0'$, being already normal to the rim at Q_0', must be the common normal at Q_0' and Q_0 if the latter is to determine a stationary value of ρ_0. The end points of the major and minor axes of an ellipse provide examples of this property.

The lines joining such point pairs as Q_0, Q_0', where $Q_0 Q_0'$ is the common normal to the rim at Q_0, Q_0', will be defined as the 'principal axes of the aperture'.

Consider now one of these point pairs Q_0, Q_0', with the above property, and let the distance $Q_0 Q_0'$ be ρ_1. A contribution to the integral is provided by the neighbourhoods of both Q_0 and Q_0'. $\theta_0' = \pi + \theta_0$, and $\delta = 0$ at Q_0 and Q_0':

$$\int \left[\frac{\cos\theta_0 \cos\theta_0'}{|\cos\delta| \, |\chi_0'|^{1/2}} \frac{\varepsilon^{-jk\rho_0}}{\sqrt{\rho_0}} + 0\left(\frac{1}{k^{1/2}}\right) \right] ds_0$$

$$\simeq -\frac{\sqrt{(2\pi)}}{k^{1/2}} \cos^2\theta_0 \frac{\varepsilon^{-jk\rho_1}}{\sqrt{\rho_1}} \left\{ \frac{\exp\left[-j\frac{\pi}{4} \text{sgn} \, \rho_0''(s_0) \right]}{|\chi_0'|^{1/2} |\rho_0''(s_0)|^{1/2}} \right.$$

$$\left. + \frac{\exp\left[-j\frac{\pi}{4} \text{sgn} \, \rho_0''(s_0') \right]}{|\chi_0|^{1/2} |\rho_0''(s_0')|^{1/2}} \right\} + 0\left(\frac{1}{k}\right) \quad . \quad . \quad . \quad (37)$$

The order of the neglected terms becomes $1/k$, since the terms $0(1/k^{1/2})$ in the integrand also contain the factor $(1/\sqrt{\rho_0})\varepsilon^{-jk\rho_0}$.

$$\left. \begin{array}{l} \chi_0 \equiv 1 - \rho_1 K_0 \\ \chi_0' \equiv 1 - \rho_1 K_0' \end{array} \right\} \qquad . \quad . \quad . \quad . \quad (38)$$

$$\left. \begin{array}{l} \rho_0''(s_0) \equiv \dfrac{d^2\rho_0(s)}{ds^2} \Big]_{s=s_0} \\[2mm] \rho_0''(s_0') \equiv \dfrac{d^2\rho_0(s)}{ds^2} \Big]_{s=s'} \end{array} \right\} \qquad . \quad . \quad . \quad . \quad (39)$$

where the parameter on which ρ_0 depends has been re-labelled s, rather than s_0, to avoid confusion.

When evaluating the derivatives in eqns. (39) it must be borne in mind that the positions of Q_0 and Q_0' are not independent. The calculation is made most simply by consideration of the geometrical properties of the curve defining the rim of the aperture in the immediate neighbourhoods of Q_0 and Q_0'.

Consider first $\rho_0''(s_0)$. It is sufficient to approximate the curve defining the rim near these points by arcs of the circles of curvature at Q_0 and Q_0', as indicated in Fig. 3. The errors
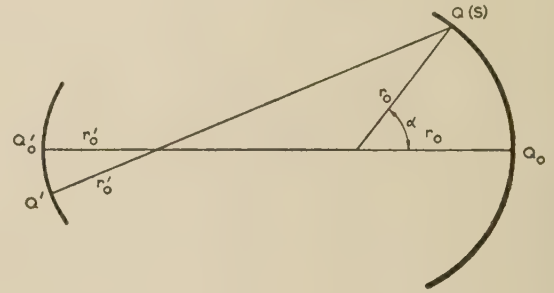

Fig. 3.—Approximation to curve defining aperture rim near ends of principal axis $Q_0 Q_0'$.

$Q(s)Q'$ is perpendicular to the circular arc at Q'.

thereby introduced then vanish in the limit as the arc length QQ_0 approaches zero.

The radii of curvature at Q_0, Q_0', will be denoted by r_0 and r_0', respectively. The line QQ' is normal to the arc at Q'; $Q_0'Q_0 = \rho_1$ and $Q'Q = \rho_0$.

Then

$$(\rho_0 - r_0')^2 = r_0^2 + (\rho_1 - r_0' - r_0)^2 + 2r_0(\rho_1 - r_0' - r_0)\cos\alpha \qquad . \quad . \quad . \quad . \quad (40)$$

Differentiation of eqn. (40) twice with respect to s, together with the aid of the relations $r_0\alpha = s - s_0$, $d\alpha/ds = 1/r_0$, then gives the following expression for $\rho_0''(s)$ when $s = s_0$:

$$\rho_0''(s_0) = \frac{\rho_1 K_0 K_0' - K_0 - K_0'}{1 - \rho_1 K_0'} \qquad . \quad . \quad . \quad (41)$$

Similarly,

$$\rho_0''(s_0') = \frac{\rho_1 K_0 K_0' - K_0 - K_0'}{1 - \rho_1 K_0} \qquad . \quad . \quad . \quad (42)$$

The denominators in eqns. (41), (42) are assumed to be non-vanishing.

From eqns. (35), (37), (41) and (42), t' is now readily obtained.

$$t' \simeq \frac{2}{k^2 A} \frac{\cos^2\theta_0}{|\Psi|^{1/2}} \mathcal{I} \left(\frac{\varepsilon^{-jk\rho_1}}{\sqrt{\rho_1}} \left\{ \exp\left[-j\frac{\pi}{4}\left(\text{sgn}\,\chi_0' + \text{sgn}\,\Phi_0\right) \right] \right.\right.$$

$$\left.\left. + \exp\left[-j\frac{\pi}{4}\left(\text{sgn}\,\chi_0 + \text{sgn}\,\Phi_0'\right) \right] \right\} \right) + 0\left(\frac{1}{k^{5/2}}\right) \quad (43)$$

where

$$\Psi \equiv \rho_1 K_0 K_0' - K_0 - K_0' \qquad . \quad . \quad . \quad . \quad (44)$$

$$\left. \begin{array}{l} \Phi_0 \equiv \rho_0''(s_0) \\ \Phi_0' \equiv \rho_0''(s_0') \end{array} \right\} \qquad . \quad . \quad . \quad . \quad (45)$$

In general, t' will consist of a sum of similar terms, each depending on the length of the corresponding principal axis, and the curvatures of the rim at Q_0 and Q_0'.

(5) THE ELLIPTICAL APERTURE

The diffraction of a normally incident plane wave by an elliptical aperture will now be examined in some detail, and the deductions will be compared with experiment. Experimental data appear to be available only for apertures of circular and elliptical form.

Consider then an elliptical aperture with major and minor axes of lengths $2a$ and $2b$, respectively, oriented so that the major axis makes an angle ϕ_1, with the positive x-axis (see Fig. 4). The incident electric field is again polarized parallel to the y-axis, and the origin of co-ordinates is taken to be at the centre of the ellipse.

(5.1) The Case of e not too Close to Zero, $1/\sqrt{2}$, or Unity

If the eccentricity is sufficiently large, the expression (43) for t'_e, the interaction contribution to the transmission coefficient, is applicable. There will be two terms in the series for t'_e, corresponding to the two principal axes of the ellipse of lengths $2a$ and $2b$.

A slight change in notation now proves convenient. The subscripts a, b, will be employed in place of 0 and 1, to denote quantities evaluated at the ends of the major and minor axes, respectively.

Then

$$\left.\begin{array}{l} \rho_a = 2a \\[4pt] \rho_b = 2b \\[4pt] K_a = K'_a = a/b^2 \\[4pt] K_b = K'_b = b/a^2 \\[4pt] \chi_a = \chi'_a = -\dfrac{1 + e^2}{1 - e^2} \\[8pt] \chi_b = \chi'_b = 2e^2 - 1 \\[6pt] \Phi_a = \Phi'_a = -2\dfrac{a}{b^2}\dfrac{e^2}{(1 + e^2)} \\[10pt] \Phi_b = \Phi'_b = \dfrac{2e^2}{b}\dfrac{(1 - e^2)}{(1 - 2e^2)} \\[10pt] \theta_a = \phi_1 \\[6pt] \theta_b = \dfrac{\pi}{2} + \phi_1 \\[8pt] A = \pi ab \\[4pt] e^2 = 1 - (b/a)^2 \end{array}\right\} \quad \cdots \quad (46)$$

Hence

$$\left.\begin{array}{l} \operatorname{sgn}\chi'_a + \operatorname{sgn}\Phi_a = \operatorname{sgn}\chi_a + \operatorname{sgn}\Phi'_a = -2 \\[6pt] \operatorname{sgn}\chi'_b + \operatorname{sgn}\Phi_b = \operatorname{sgn}\chi_b + \operatorname{sgn}\Phi'_b = 0 \end{array}\right\} \quad (47)$$

and

$$t'_e \simeq -\frac{2}{\pi e}\left[\sin^2\phi_1\frac{\sin 2kb}{(kb)^2} - \frac{b}{a}\cos^2\phi_1\frac{\cos 2ka}{(ka)^2}\right] + 0\left(\frac{1}{k^{5/2}}\right) \quad (48)$$

Adding to t'_e the geometrical-optic coefficient of unity gives finally the following expression for the transmission coefficient:

$$t_e \simeq 1 - \frac{2}{\pi e}\left[\sin^2\phi_1\frac{\sin 2kb}{(kb)^2} - \frac{b}{a}\cos^2\phi_1\frac{\cos 2ka}{(ka)^2}\right] + 0\left(\frac{1}{k^{5/2}}\right) \quad (49)$$

As might be expected, this expression becomes singular as e approaches zero, since here the stationary-phase arguments fail.

Eqn. (49) fails to yield the asymptotic form of the transmission coefficient per unit length of an infinite slit, deduced by Karp
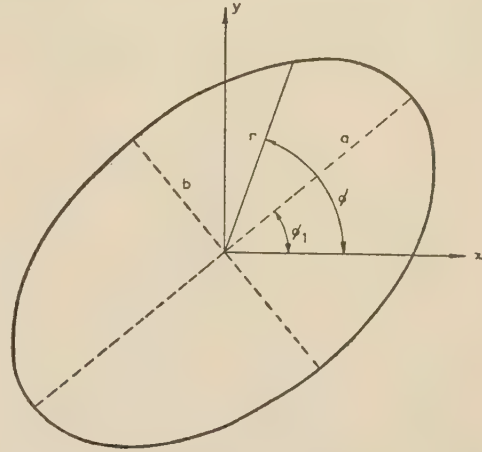

Fig. 4.—Elliptical aperture, with associated symbols.

and Russek,[8] and Clemmow,[9] in the limit as $a \to \infty$ for fixed b (i.e. as $e \to 1$). For $\phi_1 = \pi/2$ these authors found $t \simeq 1 - 1/[\sqrt{\pi}(kb)^{3/2}]\sin(2kb + \pi/4)$, where $2b$ is the slit width.

The explanation for this apparent failure lies in the fact that the slit problem is 2-dimensional, the slit length being infinite, while an assumption implicit in the present analysis is that the aperture is bounded.

The transmission coefficient for the slit may be readily obtained by slight modifications to the present method. These effectively transform the analysis into that of the first approximation of the theory of Clemmow. The plane wave E^p is still given correctly by eqns. (18), with $K'_0 = 0$. The induced current density, now independent of Y, is also given by eqns. (27), with $\delta = \theta_0 = 0$, $\theta'_0 = \pi$. The integration over Y in eqns. (29) may be immediately performed, reducing the factor $(1/R)\varepsilon^{-jkR}$ to a Hankel function. The factor $(1/z)\varepsilon^{-jkz}$ in eqns. (28 and 31) is therefore replaced by one proportional to $(1/\sqrt{z})\varepsilon^{-jkz}$. In all other respects the analysis is unchanged, and the above expression for t follows.

Although not evident from eqn. (49), the analysis is also inadequate when e is of the order of $1/\sqrt{2}$. The modifications necessary are considered in the next Section.

(5.2) The Case of $e \simeq 1/\sqrt{2}$

When the eccentricity is of the order of $1/\sqrt{2}$, special care is necessary, since χ_b vanishes and Φ_b is singular at $e = 1/\sqrt{2}$. It is readily seen that in this case, $K_b \simeq 1/(2b)$, and thus the circle of curvature of the rim at one end of the minor axis has its centre lying very close to the opposite end. In the neighbourhood of this point the interaction field is focused, since the contributions from a relatively large portion of the opposite rim arrive in phase. Focusing will also occur to a certain extent in the neighbourhood of all points on the locus of the centre of curvature of the ellipse, but only when this locus lies near the ellipse itself will the transmission coefficient be significantly affected.

To study the effects of focusing, the foregoing analysis must be repeated with the assumption that $e \simeq 1/\sqrt{2}$. Some of the steps in the analysis may be by-passed through the following considerations. Since the minor axis remains a principal axis, the integrals occurring may again be evaluated by stationary-phase arguments. Thus the first-order interaction effect on the transmission coefficient is determined solely from knowledge of the induced current density along the normal to the rim at the ends of this axis, and the nearby variation of ρ_0. Otherwise the analysis follows precisely the same lines as that of the earlier

sections. For simplicity, ϕ_1 is taken to be equal to $\pi/2$; the dependence on ϕ_1 may be reintroduced later. In particular, the term $1/(ka)^2 \cos^2 \phi_1 \cos 2ka$ will remain unchanged.
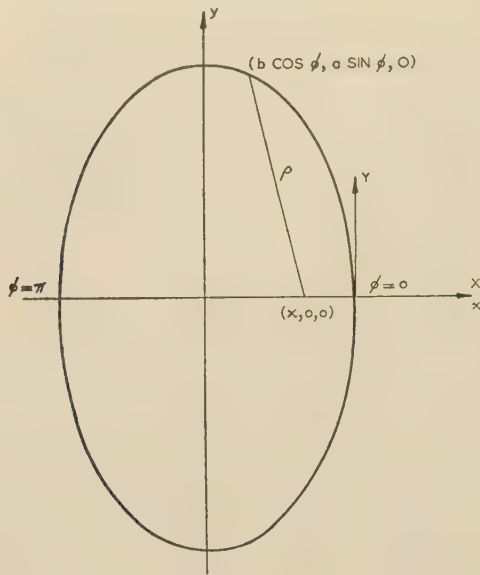

Fig. 5.—Configuration considered in discussion in which $e \simeq 1/\sqrt{2}$.

If $(x, 0, 0)$, $x > 0$, is a point on the minor axis, and ρ is its distance from $(b \cos \phi, a \sin \phi, 0)$ on the ellipse (see Fig. 5), then

$$\rho^2 = (x - b \cos \phi)^2 + a^2 \sin^2 \phi \quad . \quad . \quad . \quad (50)$$

When $\phi \simeq \pi$, ρ has approximately the following form:

$$\rho \simeq b + x + \frac{\rho''(\pi)}{2}(\phi - \pi)^2 + \frac{\rho'^v(\pi)}{24}(\phi - \pi)^4 \quad . \quad (51)$$

where

$$\left.\frac{d^2\rho}{d\phi^2}\right]_{\phi=\pi} \equiv \rho''(\pi) = \frac{a^2 - b^2 - bx}{b + x} \quad . \quad . \quad (52)$$

$$\left.\frac{d^4\rho}{d\phi^4}\right]_{\phi=\pi} \equiv \rho'^v(\pi) = \frac{bx - 4(a^2 - b^2) - 3[\rho''(\pi)]^2}{b + x} \quad (53)$$

In addition, $\qquad |\rho''(\pi)| \ll 1 \quad . \quad . \quad . \quad . \quad . \quad (54)$

$$\rho'^v(\pi) \simeq \frac{5b^2 - 4a^2}{2b} < 0 \quad . \quad . \quad . \quad (55)$$

if $e \simeq 1/\sqrt{2}$, $x \simeq b$.

The interaction electric field has a y-component only:

$$E_y^{int}(x, 0, 0) \simeq -\frac{\sqrt{2}a}{\pi}$$

$$\frac{\varepsilon^{-jk(b+x)}}{b + x} \int_0^\infty \exp\left\{-jk\left[\frac{\rho''(\pi)}{2}\phi^2 + \frac{\rho'^v(\pi)}{24}\phi^4\right]\right\}d\phi + 0\left(\frac{1}{k}\right)$$

$$. \quad . \quad . \quad . \quad (56)$$

since $\qquad ds = (b^2 \sin^2 \phi + a^2 \cos^2 \phi)^{1/2}d\phi \quad . \quad . \quad (57)$

The integral in eqn. (56) is evaluated in Section 9.2, and yields the following expression for E_y^{int}:

$$E_y^{int}(x, 0, 0) \simeq$$

$$-\frac{a}{2\sqrt{2\pi}}\left(\frac{24}{k|\rho'^v(\pi)|}\right)^{1/4}\exp\left[-j\frac{\pi}{8}\operatorname{sgn}\rho'^v(\pi)\right]\frac{\varepsilon^{-jk(b+x)}}{b + x}$$

$$\left(\Gamma\left(\frac{1}{4}\right) - jk\frac{\rho''(\pi)}{2}\left[\frac{24}{k|\rho'^v(\pi)|}\right]^{1/2}\exp\left[-j\frac{\pi}{4}\operatorname{sgn}\rho'^v(\pi)\right]\right.$$

$$\left.\Gamma\left(\frac{3}{4}\right) + 0\left\{\frac{[k\rho''(\pi)]^2}{k\rho'^v(\pi)}\right\}\right) + 0\left(\frac{1}{k}\right) \quad . \quad . \quad . \quad . \quad (58)$$

where $\Gamma(z)$ is the tabulated Gamma function.

The corresponding plane wave is readily found, namely

$$E_y^p = -E_1\varepsilon^{-jkX} \quad . \quad . \quad . \quad . \quad (59)$$

where

$$E_1 = \frac{(1 - e^2)^{-1/4}}{2\sqrt{2\pi}}\left(\frac{3}{|\zeta|}\right)^{1/4}\frac{\varepsilon^{-j2kb}}{(kb)^{1/4}}\varepsilon^{j\pi/8}$$

$$\left(\Gamma\left(\frac{1}{4}\right) + (kb)^{1/2}\frac{(2e^2 - 1)}{(1 - e^2)^{1/2}}\left(\frac{3}{|\zeta|}\right)^{1/2}\varepsilon^{-j\pi/4}\Gamma\left(\frac{3}{4}\right)\right.$$

$$\left.+ 0\left\{\frac{[k\rho''(\pi)]^2}{k\rho'^v(\pi)}\right\}\right) + 0\left(\frac{1}{k}\right) \quad (60)$$

and $\qquad \zeta = 1 - 5e^2 - \frac{3}{4}\frac{(2e^2 - 1)^2}{(1 - e^2)} \quad . \quad . \quad . \quad (61)$

The derivatives are now evaluated at $x = b$.

Eqns. (28), (29), and (31) may be combined and give for t' the following expression:

$$t' = -\frac{1}{2A}\mathscr{R}\int\int_S Z_0 J_y dS \quad . \quad . \quad . \quad (62)$$

where J_y contains the factor $\varepsilon^{-jk\rho_0}$. In the region of interest, ρ_0 takes the following form:

$$\rho_0 \simeq 2b + \frac{\Phi_b}{2}(s - s_0)^2 \quad . \quad . \quad . \quad (63)$$

with Φ_b (now in magnitude very much greater than unity) defined by eqn. (46).

Recalling that an equal contribution arises from currents near the opposite end of the minor axis, the integral in eqn. (62) may be evaluated asymptotically:

$$\int\int_S Z_0 J_y dS \sim 2\int_0^\infty Z_0 J_y(X, 0, 0)\left(1 + \frac{X}{2b}\right)dX$$

$$\left[\int_{-\infty}^\infty \exp\left(-jk\frac{\Phi_b}{2}s^2\right)ds + 0\left(\frac{1}{k}\right)\right] \quad (64)$$

where $\qquad Z_0 J_y(X, 0, 0) = -\frac{2\sqrt{2}}{\sqrt{\pi}k^{1/2}}\varepsilon^{-j\pi/4}E_1\frac{\varepsilon^{-jkX}}{\sqrt{X}} \quad . \quad . \quad (65)$

The integrations are readily accomplished. Upon simplification, the following expression is obtained for t_e':

$$t_e' \simeq -\frac{1}{\pi^{3/2}}\frac{|1 - 2e^2|^{1/2}}{e(kb)^{7/4}}\left(\frac{3}{|\zeta'|}\right)^{1/4}$$

$$\left(\Gamma\left(\frac{1}{4}\right)\sin\left(2kb - \frac{\pi}{8} + \frac{\pi}{4}\operatorname{sgn}\Phi_b\right) + (kb)^{1/2}(2e^2 - 1)\right.$$

$$\left(\frac{3}{|\zeta'|}\right)^{1/2}\Gamma\left(\frac{3}{4}\right)\sin\left(2kb + \frac{\pi}{8} + \frac{\pi}{4}\operatorname{sgn}\Phi_b\right) + 0\left\{\frac{[k\rho''(\pi)]^2}{k\rho'^v(\pi)}\right\}\right)$$

$$+ 0\left(\frac{1}{k^{9/4}}\right) \quad (66)$$

where

$$\left.\begin{array}{l} \Gamma(\tfrac{1}{4}) = 3 \cdot 625\,60 \\ \Gamma(\tfrac{3}{4}) = 1 \cdot 225\,41 \\ \zeta' = (1 - e^2)\zeta \end{array}\right\} \quad \cdots \cdots \quad (67)$$

The reintroduction of the ϕ_1-dependence gives finally

$$t_e \simeq 1 + \frac{2}{\pi e}\frac{b}{a}\cos^2\phi_1\frac{\cos 2ka}{(ka)^2}$$

$$- \frac{\sin^2\phi_1}{\pi^{3/2}}\frac{|1 - 2e^2|^{1/2}}{e(kb)^{7/4}}\left(\frac{3}{|\zeta'|}\right)^{1/4}$$

$$\left(\Gamma\!\left(\frac{1}{4}\right)\sin\left(2kb - \frac{\pi}{8} + \frac{\pi}{4}\operatorname{sgn}\Phi_b\right) + (kb)^{1/2}(2e^2 - 1)\left(\frac{3}{|\zeta'|}\right)^{1/2}\right.$$

$$\left.\Gamma\!\left(\frac{3}{4}\right)\sin\left(2kb + \frac{\pi}{8} + \frac{\pi}{4}\operatorname{sgn}\Phi_b\right) + 0\left\{\frac{[k\rho''(\pi)]^2}{k\rho'^v(\pi)}\right\}.\right)$$

$$+ 0\left(\frac{1}{k^{9/4}}\right) \quad (68)$$

where $|2e^2 - 1| \ll 1$.

(5.3) The Case of $e \simeq 0$

When the eccentricity of the ellipse is very small, evaluation of the integral in eqn. (35) by the method of stationary phase is no longer possible. Nevertheless, the integration may be effected by expansion of the quantities concerned as power series in e^2, and retention of only the lowest-order terms. The result appears to be of little practical importance, but indicates the first-order deviation from the corresponding expression for a circular aperture.

In terms of the polar co-ordinates (r, ϕ) in Fig. 4, the ellipse is represented by the following equation:

$$r = b[1 - e^2\cos^2(\phi - \phi_1)]^{-1/2} \quad \cdots \quad (69)$$

or

$$r = b\left[1 + \frac{e^2}{2}\cos^2(\phi - \phi_1) + 0(e^4)\right] \quad \cdots \quad (70)$$

If ρ is the distance between the fixed point (r_0, ϕ_0) and (r, ϕ), both on the ellipse, the condition for ρ to be stationary is obtained:

$$\frac{d\rho}{d\phi} = 0 \text{ if}$$

$$(2 + e^2[\cos^2(\phi_0 - \phi_1) + \cos^2(\phi - \phi_0)])\sin(\phi - \phi_0)$$
$$= e^2\sin 2(\phi - \phi_1)[1 - \cos(\phi - \phi_0)] + 0(e^4) \quad (71)$$

Putting

$$\phi = \phi_0 + \pi + \alpha e^2 + 0(e^4) \quad \cdots \cdots \quad (72)$$

the solution, $\phi = \phi_0'$, of eqn. (71), correct to order e^2, may be found, namely

$$\alpha = -\sin 2(\phi_0 - \phi_1) \quad \cdots \cdots \quad (73)$$

and

$$\phi_0' = \phi_0 + \pi - e^2\sin 2(\phi_0 - \phi_1) + 0(e^4) \quad \cdot \quad (74)$$

Hence to each point (r_0, ϕ_0) on the ellipse there corresponds one point (r_0', ϕ_0') at which ρ is stationary.

If, as before, $\theta_0, (\theta_0')$ is the angle between the incident electric vector and the tangent to the ellipse at ϕ_0, (ϕ_0') (see Fig. 6), then

$$\left.\begin{array}{l} \theta_0 + \beta = \phi_0 \\ \theta_0' + \beta' = \phi_0' \end{array}\right\} \quad \cdots \cdots \quad (75)$$
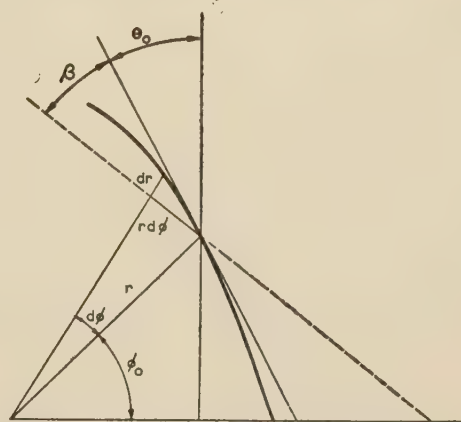
Fig. 6.—Diagram with aid of which eqns. (76) may be derived.

where

$$\left.\begin{array}{l} \tan\beta = \left[\dfrac{1}{r}\dfrac{dr}{d\phi}\right]_{\phi=\phi_0} \\[2mm] \tan\beta' = \left[\dfrac{1}{r}\dfrac{dr}{d\phi}\right]_{\phi=\phi_0'} \end{array}\right\} \quad \cdots \cdots \quad (76)$$

Then

$$\beta = -\frac{e^2}{2}\sin 2(\phi_0 - \phi_1) + 0(e^4) \quad \cdots \quad (77)$$

$$\theta_0 = \phi_0 + \frac{e^2}{2}\sin 2(\phi_0 - \phi_1) + 0(e^4) \quad \cdots \quad (78)$$

and

$$\theta_0' = \phi_0 + \pi - \frac{e^2}{2}\sin 2(\phi_0 - \phi_1) + 0(e^4) \quad \cdot \quad (79)$$

From eqn. (22),

$$\delta = e^2\sin 2(\phi_0 - \phi_1) + 0(e^4) \quad \cdots \quad (80)$$

so that

$$\cos\delta = 1 + 0(e^4) \quad \cdots \cdots \quad (81)$$

The curvature of the ellipse at (r, ϕ) is given by the following expression:

$$K(r, \phi) = \frac{2(r')^2 - rr'' + r^2}{[(r')^2 + r^2]^{3/2}} \quad \cdots \quad (82)$$

where

$$\left.\begin{array}{l} r' \equiv \dfrac{dr}{d\phi} \\[2mm] r'' \equiv \dfrac{d^2r}{d\phi^2} \end{array}\right\} \quad \cdots \cdots \quad (83)$$

Then

$$K_0' = \frac{1}{b}\left[1 - e^2 + \tfrac{3}{2}e^2\cos^2(\phi_0 - \phi_1) + 0(e^4)\right] \quad \cdot \quad (84)$$

$$\chi_0' = -1 - 2e^2\cos 2(\phi_0 - \phi_1) + 0(e^4) \quad \cdots \quad (85)$$

$$\rho_0 = 2b\left[1 + \frac{e^2}{2}\cos^2(\phi_0 - \phi_1) + 0(e^4)\right] \quad \cdots \quad (86)$$

$$ds_0 = b\left[1 + \frac{e^2}{2}\cos^2(\phi_0 - \phi_1) + 0(e^4)\right]d\phi_0 \quad \cdot \quad (87)$$

Insertion of these relations into eqn. (35), and use of the Bessel-function representation

$$J_n(x) = \frac{(-)^n}{2\pi j^n}\int_0^{2\pi}\varepsilon^{-jx\cos\beta - jn\beta}\,d\beta \quad \cdots \quad (88)$$

yields finally the following expression for t_e:

$$t_e \simeq 1 - \frac{(1-e^2)^{1/2}}{\sqrt{\pi}(kb)^{3/2}}$$

$$\left[\left(J_0(c) + \frac{e^2}{8}\left\{J_0(c) + \frac{7}{2}\cos 2\phi_1[J_2(c) - J_0(c)]\right\}\right)\right.$$

$$\sin\left[2kb\left(1 + \frac{e^2}{4}\right) - \pi/4\right]$$

$$+ \cos 2\phi_1 J_1(c) \sin\left[2kb\left(1 + \frac{e^2}{4}\right) + \pi/4\right]$$

$$\left. + 0(kbe^4)\right] + 0\left[\frac{1}{(kb)^2}\right] \quad . . . . . . \quad (89)$$

where
$$c = kbe^2/2 \quad . . . . . . \quad (90)$$

and $e \simeq 0$.

For $e = 0$, eqn. (89) reduces to the known value for a circular aperture [see eqn. (36)].

(6) DISCUSSION

The close relationship between the transmission coefficient of an aperture and the far field in the direction of propagation of the incident wave was recently exploited by Huang, Kodis and Levine[5] in an experimental determination of the transmission coefficients of circular and elliptical apertures as functions of their dimensions. The measurements on circular apertures were compared with the exact theoretical values of Andrejewski,[10] and good agreement was obtained in the range $2 \lesssim ka \lesssim 9$, where a denotes the radius of the aperture.

It should be noted that it was found possible to measure only the amplitude of the far field; the phase was determined from theoretical considerations of a variational nature which predict a phase difference between the total far-field and incident field increasing monotonically with ka from zero to $\pi/2$. For small values of ka, which are outside the scope of the present treatment, these theoretical determinations of phase are undoubtedly quite accurate, but the aperture-field distribution assumed in the variational approach is a reasonable approximation to the actual distribution only for small values of ka. It therefore seems probable that for larger values of ka the monotonic increase in phase difference will be modified in some manner. An oscillatory fine-structure (exhibited in t') is predicted by the present theory, in which, however, no attempt is made to consider the monotonic component. The latter is here assumed to be $\pi/2$ (the value predicted by geometrical optics) for all values of ka under consideration. Since for a circular aperture the agreement between experimental and exact theory was quite good, it is unlikely that a major error is introduced by the theoretically determined values of phase. In fact, if the measured far-field amplitudes (for $ka \gtrsim 6$) are employed in conjunction with a phase determined by the present theory, there is an almost negligible difference between values of the transmission coefficient calculated on this basis and the determinations of Reference 5.

Of most interest at present are the measurements performed on elliptical apertures. The ratios of minor to major axes were 1/2 and 1/3, corresponding to eccentricities of $\sqrt{3}/2 = 0.866$ and $2\sqrt{2}/3 = 0.943$, respectively. These lie in the region in which eqn. (49) is applicable. In the experimental arrangement the incident electric vector was parallel to the major axis of the ellipse, so that $\phi_1 = \pi/2$.

The transmission coefficient determined by eqn. (49) is com-
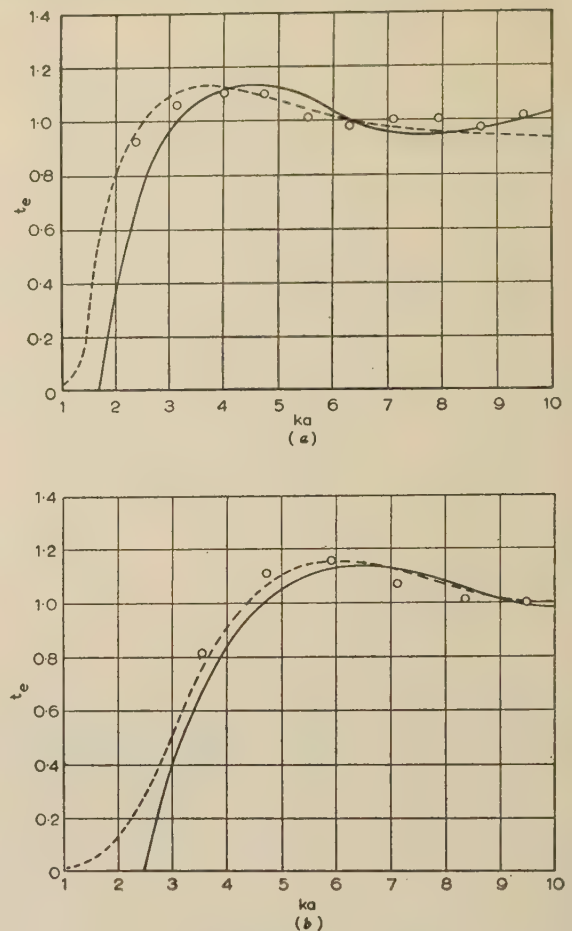
Fig. 7.—Transmission coefficient, t_e, of an elliptical aperture as a function of ka.

——— From eqn. (49) $\phi_1 = \pi/2$.
– – – – Variational approximation of Huang et al.
○ ○ Experimental measurements of Huang et al.
(a) $e = 0.866$.
(b) $e = 0.943$.

pared in Figs. 7(a) and 7(b) with the experimental measurements and with the results of a theoretical variational approximation also derived in Reference 5. Of the two cases studied, agreement of the present theory with experiment appears better when $e = 0.943$ than when $e = 0.866$. The theory then appears to predict the position of the first maximum with reasonable accuracy, and certainly compares favourably with the variational calculations.

That there is agreement at all seems rather remarkable, since the minimum dimension (i.e. length of the minor axis) of the largest aperture considered is less than two wavelengths. A fair test of the accuracy of the present theory would require measurements on apertures larger than those considered here.

A general picture of the behaviour of t_e as a function of the eccentricity may be obtained from a consideration of eqns. (49), (68) and (89). For the sake of simplicity, ϕ_1 will be taken equal to $\pi/2$. Then, as e decreases from unity towards $1/\sqrt{2}$, the behaviour is that predicted by eqn. (49). The oscillations of the form $\sin 2kb$ are damped by the factor $1/[e(kb)^2]$, resulting in an increase of amplitude, for fixed values of b, with decreasing values of e. When e, decreasing, approaches $1/\sqrt{2}$, eqn. (68) is appropriate. The oscillations take the form $\sin(2kb - 3\pi/8)$, with their amplitude decreasing towards zero because of the factor $|1 - 2e^2|^{1/2}$. Thus, in fact, the theory predicts $t_e = 1$ for all

sufficiently large values of a and b such that $e = 1/\sqrt{2}$. For $e \neq 1/\sqrt{2}$, the dependence on kb of the amplitude of oscillation becomes $1/(kb)^{7/4}$. As e decreases from $1/\sqrt{2}$, the oscillations of t_e are initially of the form $\sin(2kb + \pi/8)$, and their amplitude increases from zero. The expression (49) again becomes applicable, the amplitude of oscillation once more increasing with decrease of eccentricity. Finally, for $e \simeq 0$, eqn. (89) predicts the behaviour of t_e. The oscillations, for very small e, take the form $\sin[2kb(1 + e^2/4) - \pi/4]$, and the dependence of their amplitude on kb becomes $1/(kb)^{3/2}$.

In conclusion, it might prove worth while to discuss briefly the expected accuracy of the results that have been obtained, together with possible applications and generalizations.

The accuracy of the foregoing deductions depends ultimately on the validity of eqns. (1) as a representation of the aperture field. All other approximations that have been made appear to be quite legitimate when effects of only the first order in $1/k^{1/2}$ are considered. From the method of their derivation it is expected that eqns. (1) are asymptotic (to the first order in $1/k^{1/2}$) to the true aperture field.

There is also other evidence in support of this conclusion. As mentioned in Reference 2, the interaction component of the field in the aperture may be calculated from an assumed current distribution over the entire screen. If, at the edge of the large aperture, the singularity of the distribution is that of a half-plane, the interaction field calculated in this manner is identical, to order $1/k^{1/2}$, with that determined from eqns. (1). The solution of Chang[7] for the current density verifies the half-plane behaviour near the rim of a circular aperture. It seems hardly likely that the same is not true for large apertures of more general shape.

It may therefore be concluded that the expressions obtained for the first-order component due to interaction of the transmission coefficient should be identical with those deduced from a rigorous solution of the problem, were such a determination possible. The simplicity of this theory is such that it yields information in cases where rigorous analysis proves intractable.

It has been assumed throughout that the curve defining the aperture rim possesses positive curvature everywhere along its length. The complications which ensue when radiation from a current element on the rim to P is intercepted by an intervening portion of the screen are thus avoided. Suppose, however, that the curvature is not everywhere positive, but the aperture is such that it possesses principal axes which all lie entirely interior to the aperture. Then this complication does not arise in the regions of importance, and the expression (43) for the interaction component of the transmission coefficient is probably valid in this case also.

In addition to yielding information concerning apertures of various shapes, the formulae that have been derived might prove useful when applied to scattering by large, perfectly conducting discs, since the scattering coefficients of such discs are, by Babinet's principle, just twice the transmission coefficients of the complementary apertures. A serious limitation, however, is the present restriction to normal incidence. This could be removed by a more general formulation of the problem to oblique incidence, employing, for example, the solution to the quasi 3-dimensional half-plane problem given in Section 9.1. By methods similar to those used for grazing incidence in Section 3, the approximate current distribution on the screen when interaction is neglected could be determined, and from it the interaction component of the aperture field. The remaining analysis would follow closely that of this paper, although undoubtedly the algebra would be rather more complicated.

The aperture field is itself of some interest, and may be calculated by evaluating eqns. (1) asymptotically for a more general field point than has heretofore been considered. This aspect of the problem has been discussed with regard to a circular aperture in Reference 2, to which the reader is referred for details. In this latter case it is impossible to apply stationary-phase arguments to the evaluation of the integrals in a neighbourhood of the centre of the aperture, but for an aperture of more general shape this eventuality will not arise. The chief difficulty appears to lie in the determination of the stationary points as functions of the point P. For a circular aperture the location of these points is obvious, but even for an elliptical aperture the problem is considerably more difficult. Nevertheless, it is possible to find the solution fairly easily for the field along the major and minor axes. The numerical integration of eqns. (1) by any of the standard methods is also feasible in cases where an analytical expression is not required.

(7) ACKNOWLEDGMENTS

(8) REFERENCES

(1) MILLAR, R. F.: 'An Approximate Theory of the Diffraction of an Electromagnetic Wave by an Aperture in a Plane Screen', *Proceedings I.E.E.*, Monograph No. 152 R, October, 1955 (**103** C, p. 177).

(2) MILLAR, R. F.: 'The Diffraction of an Electromagnetic Wave by a Circular Aperture', *ibid.*, Monograph No. 196 R, September, 1956 (**104** C, p. 87).

(3) BAKER, B. B., and COPSON, E. T.: 'The Mathematical Theory of Huygens' Principle' (University Press, Oxford, 1950).

(4) LEVINE, H., and SCHWINGER, J.: 'On the Theory of Electromagnetic Wave Diffraction by an Aperture in an Infinite Plane Conducting Screen', *Communications on Pure and Applied Mathematics*, 1950, **3**, p. 355.

(5) HUANG, C., KODIS, R. D., and LEVINE, H.: 'Diffraction by Apertures', *Journal of Applied Physics*, 1955, **26**, p. 151.

(6) COPSON, E. T.: 'Diffraction by a Plane Screen', *Proceedings of the Royal Society*, A, 1950, **202**, p. 277.

(7) CHANG, H. H. C.: 'On the Diffraction of Electromagnetic Waves by a Circular Aperture' (Air Force Cambridge Research Center, Cruft Laboratory, Harvard University), Scientific Report No. 2, March, 1955.

(8) KARP, S. N., and RUSSEK, A.: 'Diffraction by a Wide Slit' (Institute of Mathematical Sciences, Division of Electromagnetic Research, New York University), Research Report No. EM-75, February, 1955.

(9) CLEMMOW, P. C.: 'Edge Currents in Diffraction Theory,' *Transactions of the Professional Group on Antennas and Propagation of the Institute of Radio Engineers*, 1956, Vol. AP-4, No. 3, p. 282.

(10) ANDREJEWSKI, W.: 'Die Beugung elektromagnetischer Wellen an der leitenden Kreisscheibe und an der kreisförmigen Öffnung im leitenden ebenen Schirm', *Zeitschrift für angewandte Physik*, 1953, **5**, p. 178.

(9) APPENDICES

(9.1) Diffraction of an Obliquely Incident Plane Wave by a Half-Plane

Let the field given by

$$E^i = (-m, l, 0)\varepsilon^{jk(lX + mY + nZ)}$$
$$Z_0 H^i = (ln, mn, -l^2 - m^2)\varepsilon^{jk(lX + mY + nZ)} \qquad (91)$$

be incident from $Z \geqslant 0$ onto the perfectly conducting screen $Z = 0$, $X \geqslant 0$. (l, m, n) are direction cosines of the incident waves, so that $l^2 + m^2 + n^2 = 1$, and time dependence, $\varepsilon^{j\omega t}(\omega > 0)$, is understood. Then the total field is given by the following expressions:

$$
\left.
\begin{aligned}
E_X &= -\left[m\Phi + m\left(\frac{2}{\pi K\rho}\right)^{1/2} \exp\left(-jK\rho - j\frac{\pi}{4}\right) \right. \\
&\qquad\qquad\qquad\qquad \left. \sin\tfrac{1}{2}\theta \sin\tfrac{1}{2}\alpha \right] \varepsilon^{jkmY} \\[4pt]
E_Y &= l\Phi\,\varepsilon^{jkmY} \\[4pt]
E_Z &= m\left(\frac{2}{\pi K\rho}\right)^{1/2} \exp\left(-jK\rho - j\frac{\pi}{4}\right) \\
&\qquad\qquad\qquad\qquad \cos\tfrac{1}{2}\theta \sin\tfrac{1}{2}\alpha\varepsilon^{jkmY} \\[4pt]
Z_0 H_X &= \left[l n\Psi + (l\lambda - m^2)\left(\frac{2}{\pi K\rho}\right)^{1/2} \exp\left(-jK\rho - j\frac{\pi}{4}\right) \right. \\
&\qquad\qquad\qquad\qquad \left. \cos\tfrac{1}{2}\theta \sin\tfrac{1}{2}\alpha \right] \varepsilon^{jkmY} \\[4pt]
Z_0 H_Y &= mn\Psi\varepsilon^{jkmY} \\[4pt]
Z_0 H_Z &= -\left[(l^2 + m^2)\Phi - (l\lambda - m^2)\left(\frac{2}{\pi K\rho}\right)^{1/2} \right. \\
&\qquad\quad \left. \exp\left(-jK\rho - j\frac{\pi}{4}\right) \sin\tfrac{1}{2}\theta \sin\tfrac{1}{2}\alpha \right] \varepsilon^{jkmY}
\end{aligned}
\right\} \quad (92)
$$

where
$$
\left.\begin{aligned}\Phi \\ \Psi\end{aligned}\right\} = \frac{1}{\sqrt{\pi}} \exp\left[jK\rho \cos(\theta - \alpha) + j\frac{\pi}{4} \right]
$$

$$
\int_{-\infty}^{\sqrt{(2K\rho)}\cos\frac{1}{2}(\theta - \alpha)} \varepsilon^{-j\tau^2}d\tau \mp \frac{1}{\sqrt{\pi}}\exp\left[jK\rho\cos(\theta + \alpha) + j\frac{\pi}{4}\right]\int_{-\infty}^{\sqrt{(2K\rho)}\cos\frac{1}{2}(\theta + \alpha)} \varepsilon^{-j\tau^2}d\tau
$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdots \quad (93)$$

and
$$
\left.\begin{aligned}
l &= \lambda\cos\alpha, \; n = \lambda\sin\alpha, \; \lambda = +\sqrt{(l^2 + n^2)} \\
K &= k\lambda, \; X = \rho\cos\theta, \; Z = \rho\sin\theta
\end{aligned}\right\} \quad (94)
$$

In the present application, $\alpha = \pi$, and the tangential component of magnetic field is given by

$$
Z_0 H_X(X, 0, 0) = \mp \left(\frac{2}{\pi k|l|X}\right)^{1/2} \exp\left(-jk|l|X - j\frac{\pi}{4}\right) \quad (95)
$$

when $X > 0$. The minus and plus signs refer to the sides of the screen on which $Z = \pm 0$ respectively.

(9.2) Evaluation of a Certain Integral

The integral concerned is of the following form:

$$
I = \int_0^\infty \exp\left[-jk(\alpha u^2 + \beta u^4)\right]du \quad \cdots \quad (96)
$$

where $|\alpha| \ll |\beta|$ and k may be taken as real.
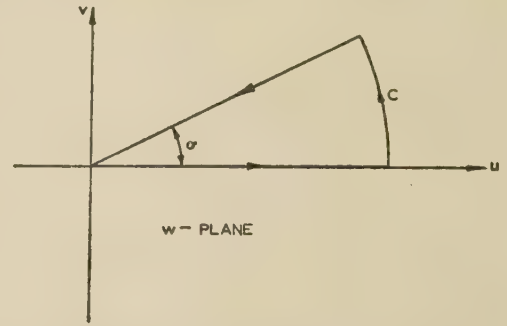
Fig. 8.—Path of integration, C, in w-plane.

Consider $\int_C \exp\left[-jk(\alpha w^2 + \beta w^4)\right]dw$, where $w = u + jv$, taken around the contour C in the w-plane illustrated in Fig. 8. There are no singularities of the integrand within or on C.

Hence $\int_C = 0$, by Cauchy's theorem.

On the circular arc, put $w = R\varepsilon^{j\phi}$. Then

$$\left|\exp\left[-jk(\alpha w^2 + \beta w^4)\right]\right| = \exp\left[k(\alpha R^2\sin 2\phi + \beta R^4\sin 4\phi)\right]$$

Choose σ equal to $-\pi/8\,\mathrm{sgn}\,\beta$. $\beta\sin 4\phi < 0$ on the arc, so $\int_{arc} \exp\left[-jk(\alpha w^2 + \beta w^4)\right]dw \to 0$ as $R \to \infty$.

Hence $\quad I = \varepsilon^{j\sigma}\int_0^\infty \exp\left(-jk\alpha R^2\varepsilon^{j2\sigma}\right)\varepsilon^{-k|\beta|R^4}dR \quad \cdots \quad (97)$

Now, if m, n, and p are all greater than zero, then

$$
\int_0^\infty x^n\varepsilon^{-px^m}dx = \frac{1}{mp^{(n+1)/m}}\int_0^\infty t^{[(n+1)/m]-1}\varepsilon^{-t}dt
$$

$$
= \frac{1}{mp^{(n+1)/m}}\Gamma\left(\frac{n+1}{m}\right) \quad \cdots \quad (98)
$$

where
$$
\Gamma(z) = \int_0^\infty t^{z-1}\varepsilon^{-t}dt \quad \cdots \cdots \quad (99)
$$

The first exponential under the sign of integration may be expanded in a power series in R^2, and the integration performed term by term, giving the following expansion:

$$
I = \exp\left(-j\frac{\pi}{8}\,\mathrm{sgn}\,\beta\right)\left\{\frac{1}{4}\frac{1}{(k|\beta|)^{1/4}}\Gamma\left(\frac{1}{4}\right)\right.
$$
$$
\left. - jk\alpha\varepsilon^{j2\sigma}\left[\frac{1}{4}\frac{1}{(k|\beta|)^{3/4}}\Gamma\left(\frac{3}{4}\right)\right] + 0\left[\frac{(k\alpha)^2}{(k|\beta|)^{5/4}}\right]\right\} \quad (100)
$$

from which eqn. (58) readily follows.

THE RELATIONSHIP BETWEEN CATHODE EMISSION, CATHODE RESISTANCE AND MUTUAL CONDUCTANCE IN RECEIVING VALVES

By M. F. HOLMES, B.Sc.

SUMMARY

The paper is divided into three parts, concerned respectively with the measurement of cathode emission and resistance and with the relationship between these two parameters and the mutual conductance of a pentode receiving valve.

Part 1 describes a method of extrapolating the zero-field total cathode emission at normal operating temperatures from the low-temperature emission measurement. It is shown that the final result is independent of contact-potential variations or potential drops across cathode, interface or collector films. The cathode work-function can also be derived from the extrapolation.

The inadequacies of existing methods of measuring cathode resistance in normal valves are considered in Part 2, which examines the possibility of substituting an empirical relation between emission and resistance whereby the average cathode resistance for a batch of valves may be quickly computed from the average cathode emission.

In Part 3 are devised relationships between the state of the cathode and mutual conductance of the valve which are based on theory and experiment. These form models which may be compared with the life-test measurements of mutual conductance and low-temperature total emission. It is found that the model which accords best with the life-test results is that which employs the concept of the cathode state being a function, not of emission alone, but of emission and resistance together.

In reaching this conclusion it is apparent that the validity of the comparison between the models and the life-test results must stand or fall by whether the assumptions which must be made in connection with the cathode work-function can be justified. With one exception it does, in fact, prove possible to justify all the assumptions required: the exception indicates, however, that the relationships which form the acceptable model have a limited, not a general, application.

LIST OF SYMBOLS

I, I_{g1} = Collector current in triodes or pentodes with control grid used as collector, mA.

I_S = Total cathode emission, amp.

$(I_s)_N$ = Total cathode emission at $1\,020°\,K$, amp.

$(I_s)_{LT}$ = Total cathode emission at $P_h = 400\,mW$, amp.

I_a = Anode current, mA.

I_∞ = Collector current at onset of retarding field conditions for cathode of infinite total emission, mA.

V = Collector voltage, volts.

V_{g1} = Applied control-grid voltage, volts.

V_0 = Value of V_{g1} at the onset of total or saturated emission, volts.

V_I = Voltage drop across any cathode-core interface, volts.

V_M = Voltage drop across the cathode matrix, volts.

V_F = Voltage drop across any collector film, volts.

V_S = Voltage drop across the vacuum at the onset of total emission, volts.

V_m = Negative minimum voltage, volts.

V_G = Equivalent diode voltage, volts.

V_a = Anode voltage, volts.

V_{g2} = Screen voltage, volts.

V_h = Heater voltage, volts.

V_c, V_{CP} = Contact potential between grid and cathode, eV.

P_h = Heater power, mW.

g_m = Mutual conductance of pentode, mA/volt.

F_1', F_2, μ_{12}' = Constants from Liebmann's paper.

α = Pentode ratio (cathode current)/(anode current).

R_k = Cathode resistance, ohms.

d, S = Cathode-collector distances, m and cm, respectively.

x_m = Distance of space-charge voltage minimum in front of cathode, m.

T = Cathode temperature, °K.

A = Cathode area, cm².

$A^{(n)}$ = Constant in emission equation (6), amp-deg^{-n}.

$\dfrac{e}{k} = \dfrac{\text{electron charge}}{\text{Boltzmann's constant}}$, $11\,606°\,K$ per volt.

q = Constant in eqn. (13).

σ_M = Matrix conductance, ohm^{-1}.

ϕ_C = Total work function of the collector, eV.

$\phi, \phi_K, \phi_1, \phi_2$ = Total work function of the cathode, eV.

ϕ_e = External work function of the cathode, eV.

$\phi^{(n)}$ = Total work function of the cathode referred to eqn. (6), eV.

Part 1. THE MEASUREMENT OF CATHODE EMISSION

(1) INTRODUCTION

It is well known that the mutual conductance of a receiving valve changes with time: at a constant anode current it deteriorates at a rate depending on the quality of the valve. It is also known that the total or saturated emission obtainable from an oxide cathode deteriorates with time and that the cathode resistance increases with time. There is general qualitative appreciation that the deterioration in emission and the increase in resistance are probably the direct cause of the falling mutual conductance. The quantitative links are not so easily determined, and this is in large measure due to the difficulties inherent in the measurement of cathode emission and resistance in a normal valve. The measurements must, of course, leave the characteristics of such a valve entirely unchanged after they have been completed. It is also impossible to make use of any artifice involving special valve construction (such as cathode probes for example) since, by definition, the measurements must be undertaken on a completely normal valve.

The paper considers the techniques available for such measurements and examines reasonable quantitative relationships between cathode emission, cathode resistance and mutual conductance.

(2) THE TWO ALTERNATIVE TECHNIQUES OF EMISSION MEASUREMENT

At normal cathode temperatures ($\sim 1\,000^\circ$ K) the measurement of zero-field total electron emission from the oxide cathodes of valves intended for space-charge-limited operation has always presented a problem. Direct measurement of normal-temperature total emission is possible only if a pulse technique is used with high collector voltages applied to the control grid. This necessitates extrapolation to zero-field conditions. An approach to the zero-field condition is important on theoretical grounds and also in practice, since the equivalent diode voltage under space-charge-limited operation is usually only of the order of $0\cdot5$ volt. Alternatively, d.c. measurements at acceptable power dissipations are practicable only at sub-normal cathode temperatures, although a closer approach is made to the zero-field requirement.

Additional difficulties arise under pulse conditions, since the measurement itself may impair the emission if the valve is not free from gas and if there are films on the collector. Both measurements are subject to doubts arising from lack of knowledge of potential distributions across the cathode and collector and from uncertainties in contact-potential variations.

Over the past few years those engaged on thermionic-valve investigations in the Post Office have found it necessary to make comparative emission measurements on large numbers of valves. It was therefore desirable that the apparatus used should be simple, permit a rapid testing rate and present no difficulty in co-ordinating agreement between different investigators. The d.c. measurement at low temperatures, requiring essentially only voltmeters, milliammeters and l.t. accumulators, is particularly suitable in these respects, and has therefore been adopted as a standard technique. Because of the difficulty of extrapolating from low to normal temperatures, measurement at the former temperature has been used by itself as a basis for comparative emission studies; some justification of this procedure is considered in Part 3.

The low-temperature technique has been described by Metson, Wagner, Holmes and Child,[3] and the basic details of the method still apply. Improvements have been made in the control of heater power and in the testing rate to such an extent that it is now practicable to measure the emission of 10 valves in about 5 min.

The difficulties of the double extrapolation to zero field and normal temperature are considered below, and a more elaborate technique is described whereby extrapolation may be effected. Greater reliability and additional data are, of course, obtained only at the expense of a slower testing rate, and it is as a supplement to, and a check of, the standard method in isolated cases that the new method will find its greatest use.

(3) MORE DETAILED EXAMINATION OF THE LOW-TEMPERATURE MEASUREMENT

The measurement described in the earlier paper[3] is undertaken at one heater power (400 mW for a 2-watt heater) corresponding to a cathode temperature of about 700° K, and the collector voltage applied to the control grid is kept constant at 5 volts. These conditions correspond to measurement at a single point, P, on a typical log I_{g1}/V_{g1} characteristic, as shown in Fig. 1. A family of curves exists for any one cathode, each being similar to that shown in Fig. 1 and identified by a particular cathode temperature.

It is useful to identify the component parts of the log I_{g1}/V_{g1} curve with the potential diagrams of the cathode-collector system. The curve in Fig. 1 comprises three parts, AB, BC and CD. AB corresponds to retarding field conditions with no
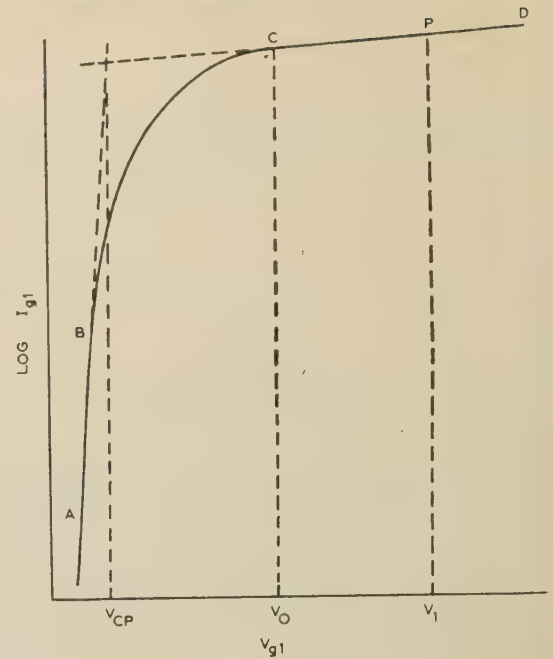
Fig. 1.—Typical log I_{g1}/V_{g1} characteristic of a modern receiving valve ($T \sim 700^\circ$ K).

potential minimum existing between cathode and collector surfaces; under the conditions corresponding to B a potential minimum appears at the surface of the collector, giving a zero potential gradient or zero field at that surface. BC corresponds to space-charge-limited conditions with the potential minimum occurring in the space between collector and cathode; at C there occurs the onset of total or saturated emission as the potential minimum disappears into the surface of the cathode and zero-field conditions apply. The part CD corresponds to Schottky accelerating-field conditions, once again without a potential minimum between cathode and collector.

It is clear that the condition represented by the point C is the one which is of most interest for zero-field emission, and the relevant potential diagram is shown in Fig. 2. Here the Fermi
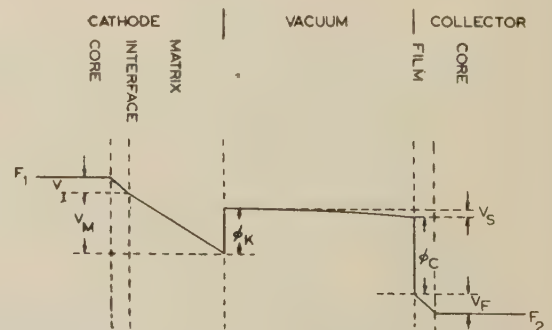

Fig. 2.—Voltage diagram of an oxide-cathode diode at onset of saturation (Point C, Fig. 1).

level, F_1, of the cathode core is used as a reference and positive voltages are measured downwards. The voltage difference between the Fermi levels F_1 and F_2 of the cathode and collector cores corresponds to the voltage applied externally to the cathode-collector system, i.e. to V_0 in Fig. 1, with the collector positive with respect to the cathode. It can be seen from Fig. 2 that

$$V_0 = V_I + V_M - \phi_K + V_S + \phi_C + V_F \quad . \quad . \quad (1)$$

Now the contact potential between the cathode and the collector is given by

$$V_{CP} = \phi_C - \phi_K$$

taken here as positive if $\phi_C > \phi_K$. Thus

$$V_0 = V_I + V_M + V_S + V_F + V_{CP} \quad . \quad . \quad . \quad (2)$$

the signs of all terms being positive if $\phi_C > \phi_K$.

Consequently, when measuring total emission from a cathode under conditions corresponding to the point P in Fig. 1, at a constant applied collector voltage, the emission is subject to a field effect dependent in some way upon the voltage difference $V_I - V_0$, and therefore upon V_0. If several measurements of emission are taken under the condition P at different low temperatures, and are used to extrapolate to the emission at normal temperatures, then it is implied that either V_0 is invariant with temperature or the field effect is negligible. It will be shown in Section 5 that the field effect is certainly not negligible, and, through examination of the temperature dependence of the terms on the right-hand side of eqn. (2), it will be shown in the next Section that V_0 is not independent of temperature. Consequently, the extrapolation described is invalid.

(4) TEMPERATURE VARIATION OF THE FIELD EFFECT

The terms on the right-hand side of eqn. (2) have different degrees of temperature dependence. It is reasonable to assume, for instance, that V_{CP} is substantially constant with temperature over a range of about $150° K$ when p.d.'s which affect the issue are of the order of $0·1$ volt.

For the voltage drop across the cathode matrix under saturation conditions the relation

$$V_M = I_S/\sigma_M \quad . \quad . \quad . \quad . \quad . \quad (3)$$

holds, so that

$$V_M \propto \varepsilon^{-e\phi_e/kT}$$
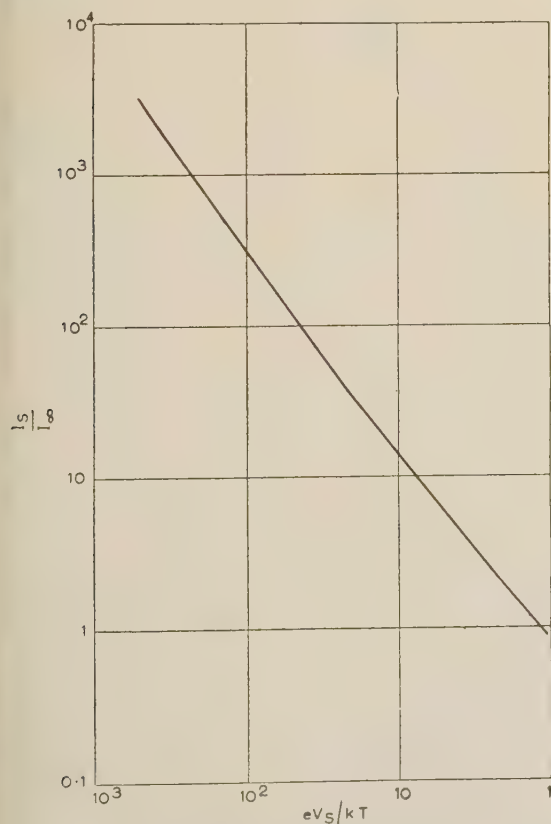
Fig. 3.—General curve relating emission, I_s, and cathode-collector voltage, V_S, at onset of saturation.

Under these circumstances there is a definite dependence of V_M on temperature.

A relation similar to eqn. (3) holds for V_I and V_F if an interface layer or a collector film is present. In these cases, however, I_S still refers to the cathode emission current, whereas the conductance term refers to either the interface or the film. It is unlikely that the total work function of the cathode will equal the internal work function of either the interface or the film—the only condition which would give temperature independence. It is therefore probable that V_I, V_F and V_M are temperature dependent, although it would not be easy to determine experimentally the quantitative nature of the dependence.

It remains to consider the voltage drop between the cathode and the collector surfaces which produces a potential minimum at the former surface, and here a theoretical determination of the temperature dependence can be attempted. The magnitude of V_S can be derived from Langmuir's equation relating space-charge-limited current with the cathode-collector voltage in a diode system. A simplified method of treating this problem has been developed by Ferris,[9] and using his approach it can be shown that V_S increases with temperature. Two curves are included to illustrate this change: Fig. 3 is a general curve relating $\log I_S/I_\infty$ and eV_S/kT, where I_∞ is a constant used by Ferris—a function of valve geometry and temperature which is defined as the current corresponding with the point B in Fig. 1 for a cathode of infinite emission. It is possible to derive further curves from this general curve which will relate V_S and T for any particular level of cathode emission and any diode system. Fig. 4 shows an example for the following special conditions:

(i) A total cathode emission given by the equation

$$I_S = 100T^{5/4}\varepsilon^{-e\phi/kT} \text{ amperes} . \quad . \quad . \quad . \quad (4)$$

with $\phi = 1·1 \text{ eV}$.

(ii) A diode system defined by

$$I_\infty = 0·245 \times 10^{-3}\left(\frac{T}{1\,000}\right)^{3/2}\frac{A}{S^2} \text{ milliamperes} \quad . \quad . \quad (5)$$

where A = Cathode area = $0·45 \text{ cm}^2$.
S = Cathode-collector separation = $0·015 \text{ cm}$.
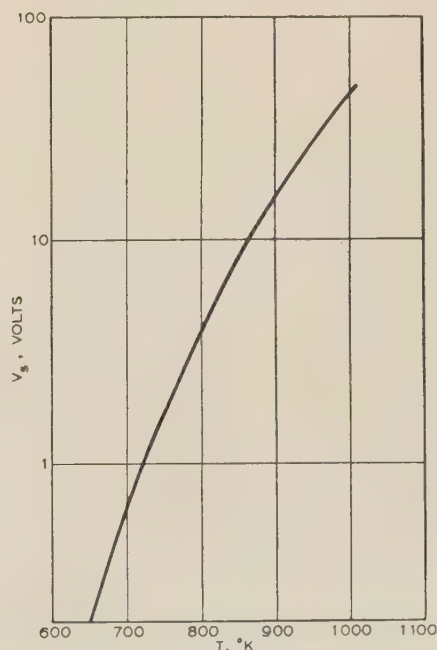
Fig. 4.—V_s/T relation for diode system defined by eqns. (4) and (5).

On the basis of this discussion it is clear that V_0 is temperature dependent. It is also clear that it is difficult to determine the theoretical value of V_0 at different temperatures, because it is so dependent on work functions which are not easy to determine experimentally on normal valves. It is also difficult to determine V_0 from the deviation of DCB in Fig. 1 from a straight line to a curve, owing to the restricted length of CD. (The voltage corresponding to the point D should not exceed 5 volts, owing to the danger of film breakdown.)

(5) AN ALTERNATIVE APPROACH TO THE FIELD CORRECTION

An alternative approach to the problem of correcting for the field effect in the low-temperature emission measurement may be seen by considering in more detail the curve obtained by recording emission measurements as a function of temperature at constant collector voltage. This curve was rejected in Section 3 as a basis for extrapolation owing to lack of means of assessing the field effect. The curve is, of course, a plot of $\log I$, the recorded emission, against $1/T$ at constant voltage, and a typical example is shown in Fig. 5 (curve SRQ) for a collector voltage of $3 \cdot 5$ volts.
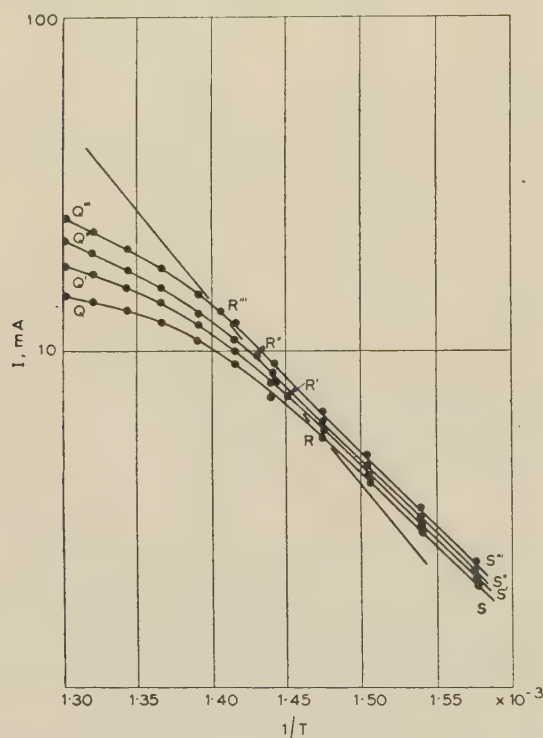
Fig. 5.—Typical $\log I/(1/T)$ characteristic with collector voltage, V_1, as a parameter.

The rejected extrapolation is the straightforward extension of SR to give an emission value at normal cathode temperature.

Referring once again to Fig. 1, it will be appreciated that P for a collector voltage of $3 \cdot 5$ volts may be in the region CD, may coincide with C or may be in the region BC, as the temperature increases. This is, in the first place, due to the fact that V_{CP} is substantially constant, as was pointed out in Section 3. The location of P in connection with the three possibilities, CD, C or BC, is therefore dependent on the magnitude of $V_0 - V_{CP}$. Now it can be shown [eqn. (2)] that $V_0 - V_{CP} = V_M + V_S$, provided that there is no interface or collector-film resistance. It can also be shown that the increase of V_S with temperature is the controlling factor when considering the

changes of $V_M + V_S$, and therefore of $V_0 - V_{CP}$ with temperature over the range 650–750° K. At very low temperatures, P will be in the region CD and the measurement will be made under accelerating field conditions. As the temperature rises, the increase of V_S will cause $V_0 - V_{CP}$ to increase, and C will move towards P. When P coincides with C the space-charge minimum will just appear at the cathode surface and will cause a zero off-cathode potential gradient. As the temperature increases further, C will pass further to the right and P will fall in the region BC, where space-charge-limitation applies.

In Fig. 5 the portion of the curve SR is identified with the location of P in CD, i.e. with accelerating field conditions, and the portion of the curve RQ is identified with the location of P in BC, i.e. with space-charge-limited conditions; the point R, where the straight line SR becomes the curve RQ, is identified with coincidence of P and C, i.e. with zero off-cathode potential gradient. Thus R can be regarded as the only point on SRQ measured under effectively zero field conditions and the emission at that point will be independent of any considerations of contact-potential variations or potential drops across the cathode, the interface or the collector film.

The points R′, R″ and R‴ are also measures of zero field emission at different temperatures, being derived from curves having collector voltages V_1', V_1'', V_1''', equal to $4 \cdot 0$, $4 \cdot 5$ and $5 \cdot 0$ volts respectively. Now all the emission equations, of which

$$I_s = A^{(n)} T^n \exp\left(-e\phi^{(n)}/kT\right) \quad \ldots \quad (6)$$

is the general form, imply a zero field. [The use of (n) as distinct from n implies an identifying superscript as distinct from a power index.] It is therefore argued that R, R′, R″ and R‴ are the only points in the $\log I_s/(1/T)$ field of Fig. 5 which may be used to extrapolate the emission to normal temperatures. If this is done, the form of emission equation implied is that due to de Boer, where $n = 0$ in eqn. (6).

As is well known, it is not possible to discriminate experimentally between the several emission equations using different values of n, such as $n = 0$ (de Boer), $n = \frac{1}{2}$ (Richardson) and $n = 5/4$ or 2 (semi-conductor theory). Here the use of $n = 0$ is taken for experimental convenience rather than in support of the de Boer equation against the rest. The value of work function so derived is $\phi^{(0)}$, but, if necessary, the points R, R′, R″ and R‴ can easily be replotted to conform with any other chosen form of index, e.g. 2 or 5/4. If this is done the values of work function are not very different, being reduced from the $\phi^{(0)}$ value by $0 \cdot 05$ to $0 \cdot 1$ eV.

Before proceeding to a more detailed consideration of the experimental method and results in the next Section, it is worth while commenting on Fig. 5, which forms the basis of this Section. First, the compensation for the field effect is seen to be of considerable importance. Extrapolation of SR to 1 000° K yields a normal-temperature emission of 700 mA, while extrapolation of RR‴ yields $2 \cdot 2$ amp. The value of $\phi^{(0)}$ based on RR‴ is $1 \cdot 08$ eV, and a work function of $0 \cdot 86$ eV may be calculated on the basis of SR, but while $1 \cdot 08$ eV is within the range of work functions generally accepted for the (BaSr)O cathode, $0 \cdot 86$ eV is not.

The linearity of the SR plot, which enables the onset of space-charge limitation to be defined at R, where curvature starts, is due to the fact that the relation between V_S and T is roughly linear over a limited temperature range. It is interesting to note that, moving to the right from R along RS, V_S decreases with decreasing temperature. There comes a time, as V_S decreases sufficiently in magnitude, when the relatively slow change of V_M with temperature supervenes. At this point the $\log I_s/(1/T)$ plot increases noticeably in slope, i.e. takes an abrupt kink

ownwards. Fortunately this occurs at a temperature between
0 and 100° K lower than the zero-field point (it occurs where
0³/T = 1·6 in Fig. 5). This permits the linearity of SR to be
stimated over a reasonable range.

(6) EXPERIMENTAL DETAILS AND RESULTS

A simple circuit such as that shown in Fig. 6 has proved to be
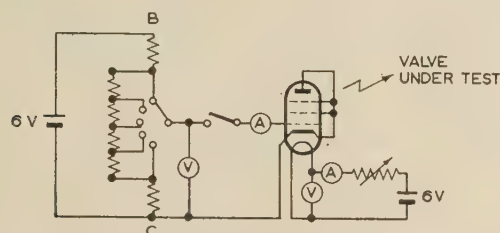ery suitable for measuring the curves shown in Fig. 5. The low

Fig. 6.—Circuit used for emission measurements.

esistance (∼6 ohms) of the voltage divider BC provides sub-
tantially constant voltage sources of 5·0, 4·5, 4·0, 3·5 and
·0 volts, although it will be realized that the precise value of
ne voltage is immaterial, provided that it remains constant.
If a single valve type is being tested, a common heater-
ower/cathode-temperature relationship can be used without
erious error. Thus the heater power is set at some particular
alue and the collector currents are recorded for the four or five
ollector voltages. Four or five points are therefore obtained on
ne log $I/(1/T)$ curves at one value of $1/T$. The readings are
epeated at different heater powers. The measurements neces-
ary to complete a set of curves can be taken in about 15 min,
llowing about 1 min for the cathode temperatures to become
tabilized before all readings except the first, where 3 min should
e allowed from cold conditions. All heater-power and
ollector-voltage changes are made with the control-grid key
pen, and the collector current readings are taken by momentary
epression of the key.
From experience the lowest values of I which need to be
ecorded, corresponding to S, S′, etc., are of the order of 1 mA
t a cathode temperature of about 640° K. The points Q, Q′,
tc., will have been reached in almost all cases before I reaches
0 mA and before T reaches 760° K. In many cases it will be
ound unnecessary to record currents higher than 10 mA.
ecause of film breakdown, the highest collector voltage used
s 5 volts. In general, about 10 sets of readings at 10–15° K
ntervals will be found adequate. The whole family of curves
s shifted to the left for valves with poor emission levels, and to
ne right for exceptionally good valves.
The distance over which RR‴ must be extrapolated is sub-
tantial if the valve is good, and is much reduced as the valve
eteriorates. Support for the reliability of the extrapolation
s derived from the results quoted in the next paragraphs. The
alues of ϕ and total emission at normal temperatures are in
greement with generally accepted results, and in addition,
ome correlation is found between results obtained using this
xtrapolation and results depending on extrapolation from pulse
measurements.
Values of $\phi^{(0)}$ [eqn. (6)] for 27 valves are plotted in the form
f a histogram in Fig. 7, together with extrapolated values of I_s
t 020° K (roughly equivalent to $V_h = 6·0$ volts). It will be
een that a mean value of $\phi^{(0)}$ of about 1·4 eV is obtained,
hich would give a value of $\phi^{(5/4)}$ of about 1·3 eV. The mean
ormal-temperature emission of about 1·5 amp corresponds to
amp/cm². The valves are of the deep-sea repeater type, and
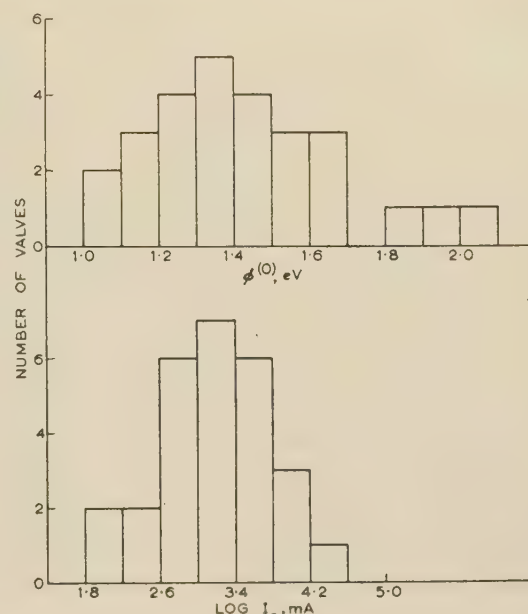
Fig. 7.—Values of $\phi^{(0)}$ and I_s at 1 020° K for 27 valves.

almost all of them have platinum cores. Within the relatively
small sample no clear correlation could be detected between
$\phi^{(0)}$ and I_s, i.e. that high values of I_s, for example, are not
uniquely associated with low values of $\phi^{(0)}$.

One other item of interest should be noted. The measure-
ments on the 27 valves were made whilst they were on life tests
and after varied periods of time. If the average work functions
are considered on a time basis, the following information can
be derived:

Valves measured		Average value of $\phi^{(0)}$ eV
Between 0 and 500 hours	..	1·3
Between 500 and 1 500 hours	..	1·4
Over 1 500 hours	..	1·6

There is thus some evidence of increasing work function with
life.

Despite the lack of a really successful pulse technique, an
attempt was made to compare the emission of some valves
measured under both pulse and low-temperature methods. For
the pulse measurement, extrapolation to zero field was made
using the Schottky line. Agreement to within about 20% was
obtained in some cases; in others the pulse-derived emission
was only 50% of the low-temperature derivation, although here
clear indication of deterioration under the influence of the high-
voltage pulse (100 volts) was obtained. It is, however, felt that
a more rigorous comparison should be undertaken using single
pulses in place of the pulse recurrence frequency of about
1 000 pulses/sec that is sometimes quoted.

Part 2. CATHODE RESISTANCE IN RECEIVING VALVES

(7) INTRODUCTION

The oxide cathode is of finite thickness and consequently has
an associated conductivity and resistance. In the cathodes of
most modern receiving valves the resistance has two components,
namely that of the oxide coating itself and that of an interface

layer between the coating and the metallic cathode core. It is not proposed to consider the interface layer in the paper, because the associated experimental work is based on valves whose cathodes do not develop interface resistances, namely valves with platinum cores. There is also a reasonable possibility of eliminating the interface entirely in future commercial valve production through the adoption of new cathode-core alloys. The technique for measuring interface resistance is, in any case, well known and is described in detail elsewhere.[1] The technique for measuring the resistance of the oxide coating itself, on the other hand, is very difficult, and a review of the problems involved is given in the next Section.

The resistance associated with the oxide cathode is not, so far as the paper is concerned, equated to any physical model. It has been suggested by Loosjes and Vink[2] that the mechanism of conduction through the coating consists of a combination of pore conduction and solid conduction, with the former predominating at normal cathode temperatures ($\sim 1\,000^\circ$ K). Most investigators favour this model, but the model requiring solid conduction alone has not been entirely excluded. Although a detailed physical model is not required here, it will be seen that certain properties of the oxide-cathode resistance are stressed by the results obtained in later Sections.

(8) THE MEASUREMENT OF CATHODE RESISTANCE

The measurement of cathode resistance on normal receiving valves is subject to the same limitations that apply to the measurement of cathode emission. The measurement must leave the characteristics of the valve unchanged after completion and must be undertaken without the use of any special valve construction. This second limitation immediately excludes the only really reliable methods of resistance measurement, namely the use of embedded probes or measurements on cathode blocks between large electrodes.

Despite these limitations, attempts have been made to measure cathode resistance in normal valves. Two methods have been suggested which are based on the shape of the current/voltage characteristic, using the control grid as collector, one being based on the displacement[3] of the characteristic and the other on the curvature[4] of the characteristic in the retarding field region. Both methods suffer from the disadvantage that the derived resistance includes the resistance of any film present on the collector, and both give results which vary with cathode current, i.e. they vary with that portion of the current/voltage characteristic used to make the calculation. The reason for this is not completely understood, and the quantitative basis for the variation is lacking. Thus it is not considered profitable to use this type of measurement at this stage, although information gained in such measurements is used in framing a hypothesis to test against experimental evidence.

In the discussion on Reference 3 Eaglesfield suggested that measuring the shift in the grid-voltage/grid-current characteristic when screen-grid current is caused to flow by an applied voltage was the simplest method which might be applied to a normal valve for the derivation of cathode resistance. The results yielded are of the right order, but are subject to variations depending on the grid current used to measure the shift of the characteristic and on the screen-grid voltage. This lack of precision is fundamental to the method and has prevented its use here. Finally, Nergaard[5] has suggested an equivalent circuit which may be employed in a bridge network in order to determine the values of R and C to be associated with the oxide cathode, against which the equivalent circuit is balanced. This type of method has been criticized by Tillman, Butterworth and Warren[6] in their work on the dependence of mutual conductance on

frequency. They hold that, even when considering the interface impedance on its own, its representation as two elementary components only is inadequate for the explanation of certain experimental results. It is consequently felt that the combined representation of oxide cathode and interface by the networks suggested by Nergaard may require qualification.

At the moment, all available methods of measuring the resistance of the oxide cathodes of normal receiving valves require some form of qualification, and thus none of these methods will be used directly in the examination of the relationship between emission, resistance and mutual conductance. As an alternative approach it is proposed to calculate the resistance of the cathode from the emission in a manner which has some experimental justification. The validity of the calculation will then be examined against other critical experimental data.

(9) CALCULATION OF CATHODE RESISTANCE

In order to calculate the resistance of the cathode, use is made of the proportionality of emission and conductance reported by Metson et al.[3] and by Hannay et al.[10] A comparison was made between the resistances of oxide cathodes measured by Dalman's method and by the method used in Reference 3. The results were of the same order in both cases, and because of the mutual support it was decided to use the proportionality equation [eqn. (8)] from Reference 3 in this investigation. The equation is more conveniently expressed as

$$I_s R_k = 1\cdot 6 \qquad \ldots \ldots \quad (7)$$

Although this equation was established at low cathode temperatures (700–800° K), it is intended to extend its use up to normal cathode temperatures ($\sim 1\,000^\circ$ K).

A corresponding equation can be deduced from Hannay, MacNair and White's paper[10] with a figure of $0\cdot 5$ in place of the $1\cdot 6$ used in eqn. (7). Additional comment on the choice of $1\cdot 6$ will be given in Section 13.

When considering the relationship between emission, resistance and mutual conductance in the Sections which follow, eqn. (7) will be used, with other factors, to derive relationships between the state of the cathode and the mutual conductance of the valve. When using eqn. (7) in any context the emission and resistance must, of course, be measured at the same temperatures. If the equation is applied at normal temperatures, it is the normal-temperature total emission which is involved. This may be derived from the method described in Part 1, but, if many valves are to be subject to examination, this procedure would consume excessive time. Alternatively, an approximate calculation of the normal-temperature emission can be made by using the equation

$$(I_s)_N = 1\,000(I_s)_{LT} . \qquad \ldots \ldots \quad (8)$$

This relation can, of course, best be used when considering the average results for batches of valves, rather than in individual cases. It can be justified to some extent for batch treatment in that the multiplying factor of $1\,000$ used in eqn. (8) assumes a value of $\phi^{(0)}$ of $1\cdot 32$, which is in reasonable agreement with the experimentally determined average value of $\phi^{(0)}$ taken from Fig. 7. It is appreciated that the use of eqn. (8) implies a constant cathode work-function, not only from valve to valve, but also with life, and that one conclusion reached in Part 1 was that there was some evidence of an increasing work function with life. It is nevertheless intended to use the equation and, by comparison with experiment, to observe when it breaks down, so obtaining additional information on the trend of change of work function. Further consideration of this trend will be deferred to Section 13.

Before closing this Section a final comment on the use of

eqn. (7) is necessary. The general objections to the method of resistance measurement used in deriving this equation were stated in the second paragraph of Section 8. The first objection, the inclusion of any unwanted collector-film resistance, was eliminated in the data used to establish eqn. (7). This was made possible in unpublished work by Child, who confirmed the absence of such films in the particular valves involved, using a 2-frequency method. The second objection, the variation of computed cathode resistance with cathode current, must stand. Here the use of the equation can be justified only through comparison of the results obtained from its application with other experimental evidence. This comparison will be made in Section 12.

Part 3. THE RELATIONSHIP BETWEEN CATHODE EMISSION, CATHODE RESISTANCE AND MUTUAL CONDUCTANCE

(10) INTRODUCTION

Before considering the relationship between cathode emission, cathode resistance and the mutual conductance of the valve, it is necessary to refer briefly to the experimental derivation of the mutual conductance. All the measurements reported here were made with an a.c. bridge method at a frequency of 1 kc/s with a control-grid signal amplitude of about 50 mV, d.a.p. The repeatable accuracy of the bridge is considered to be about 0·1%; the absolute accuracy is good, but is, of course, dependent on the accuracy of the meters (among other components) used.

The following plan will be used to study the interdependence of these parameters: First, it is possible to calculate a relationship between the total emission of the cathode and the mutual conductance for a particular valve structure, and the manner of doing this is described in Section 11. Second, it is possible to take the calculation a stage further and allow for the effect of cathode resistance. We then have two relationships, the first between cathode state, judged on emission alone, and mutual conductance, and the second between cathode state, judged on emission and resistance, and mutual conductance. These form two models which may be set against experimental data to see which provides the closer fit. It will be shown that the large-scale life tests undertaken at the Post Office Research Station provide a block of experimental data which can well be used in assessing the validity of the two models.

Before leaving this introduction it is worth while to give some details of the life-test procedure. The valves concerned are pentodes of the Post Office types 6P10 and 6P12. The first is very similar to the CV 138, and the second is almost identical, except that the screen grid has been moved closer to the control grid in order to permit operation with $V_{g2} = 60$ volts, instead of the 250 volts required for the 6P10. Both types are run under constant-cathode-current conditions, achieved by high d.c. feedback (6 kilohms in the cathode lead offset by a stabilized positive potential of 50 volts applied to the control grid). Consequently any changes in contact p.d., or other factors affecting the valve working point, have a reduced effect on the currents drawn to anode and screen. The anode and screen voltages are also stabilized and the heaters are run from a stabilized a.c. source.

(11) THE RELATIONSHIPS BETWEEN CATHODE STATE AND MUTUAL CONDUCTANCE

The basic current/voltage relationship used in any treatment of valve operation is afforded by the Langmuir equation for a planar diode:

$$I = 2·335 \times 10^{-6} \frac{(V - V_m)^{3/2}}{(d - x_m)^2} \left[1 + \frac{T^{1/2}}{40·3(V - V_m)^{1/2}} \right] \quad . \quad (9)$$

In order to apply this equation to pentodes with substantially planar cathode surfaces, it is necessary to reduce the pentode structure to the equivalent diode, and here the treatment of the problem given by Liebmann[7] is followed. In particular, use is made of the cathode sectionalization described by Liebmann to allow for the variable-mu effect.

Since the total cathode emission, I_s, occurs in the derivation of V_m and x_m in Langmuir's equation, it is possible to construct a relationship between the emission and the mutual conductance of the pentode. A set of equivalent diode voltages, V_G, were calculated for a particular pentode structure, for particular values of screen voltage, V_{g2}, and contact potential, V_c, between the grid and the cathode, and for a series of control-grid voltages; Liebmann's equation

$$V_G = F_1'\left(V_{g1} + V_c + \frac{F_2 V_{g2}}{\mu_{12}'}\right) \quad . \quad . \quad . \quad (10)$$

was used for this purpose, where the factors F_1', F_2 and μ_{12}' may be calculated from the valve geometry.[7] These values of V_G were then applied to Langmuir's equation as V, the collector voltage, to calculate the current/voltage relation for the equivalent diode at any chosen value of I_s. The computed values of equivalent diode current were then plotted against the value of V_{g1} used to compute V_G, so producing the cathode-current/grid-voltage characteristic of the pentode. This characteristic was transformed to the anode-current/grid-voltage curve through the application of a suitable factor, α. Further curves of this type were calculated for different values of I_s, and from this set of curves it is possible to derive another family with mutual conductance plotted against I_s, having constant I_a as parameter. This family of curves provides the first model for the relationship between cathode state, as measured by emission alone, and mutual conductance.

In this connection it should be remembered that no assumption is made in Langmuir's equation as to the cause of different values of I_s. The calculation is developed from Child's law, where the initial velocities of the electrons are neglected, by making use of Maxwell's velocity distribution of the emerging electrons. Here, however, it is necessary to consider one aspect of the nature of the change of I_s. If I_s is assumed to change owing to a change in the work function of the cathode, the appropriate correction must be made to V_c and therefore to V_G. This correction has, in fact, been made in the calculations used here. It can, however, be shown that, so far as the relation between mutual conductance (at constant anode current) and total emission is concerned, the correction is immaterial. The same result is obtained, within the accuracy of the computation used, as if the contact potential had been assumed to be constant. The major effect of the correction is in its control of the position of the I_a/V_{g1} characteristic along the grid base. If the correction is applied, the position of the characteristic remains substantially invariant with changing I_s, except for very low values of I_s, as first explained by Gysae and Wagener.[8] If the correction is not made, i.e. if I_s is assumed to change from some cause other than change of work function, the characteristics are spread over a much wider portion of the grid base. However, this difference is outside the scope of the investigation.

At a later stage in the paper it will be found necessary to use the concept of work function in order to translate low-temperature to normal-temperature total emission through use of eqn. (8) or some similar relation. This need arises in the attempt to check the validity of the model relation, developed above, against life-test records. The fact that the model is substantially independent of the concept of work function will exclude the complication of amending the model to suit the necessary work function

adopted for the translation. The relative unimportance of work function and contact potential is doubly fortunate here, for it must be appreciated that the work function of the control grid is usually unknown.

Typical examples of calculated anode-current/grid-voltage characteristics and the complementary mutual-conductance/total-emission curves are shown in Figs. 8 and 9(a), the valve used
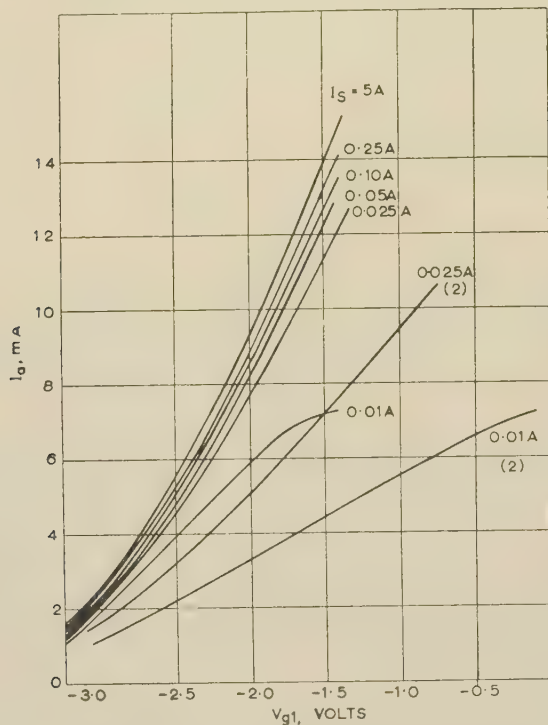

Fig. 8.—I_a/V_{g1} characteristics, with I_s as parameter, for the 6P10 valve, calculated in accordance with the first model.

The curves marked (2) show two of the characteristics modified by the inclusion of a cathode resistance inversely proportional to the emission I_s.

being the 6P10. The solution of the Langmuir equation was effected using the method developed by Ferris,[9] and a cathode temperature of $1\,020^\circ$ K was used for the calculation. The cathode-control-grid spacing is assumed to be 0.15 mm and the cathode area 0.45 cm^2, which gives $I_\infty = 0.505$ mA at $1\,020^\circ$ K [eqn. (5)]. Values of I_s used were derived from values of I_s/I_∞ of $10\,000$, $2\,000$, 500, 200, 100, 50 and 20. Curves of I_a/V_{g1} were thus calculated for a range of total emission extending from 5 to 0.01 amp. A value of 0.8 volt was assumed for V_c at $I_s/I_\infty = 10\,000$, and a value of 1.1 eV for ϕ_K, the work function of the cathode, under the same circumstances. The following equation

$$\log_\varepsilon \frac{(I_s)_1}{(I_s)_2} = \frac{e}{kT}(\phi_2 - \phi_1) \quad . \quad . \quad . \quad . \quad (11)$$

was used to calculate changes in work function necessary to compute changes in V_c with total emission. A constant value of 1.9 eV for ϕ_c, the work function of the control grid, was assumed to hold throughout the calculations; this figure is, of course, derived from the initial values of V_c and ϕ_K. The I_a/V_{g1} curves were calculated for a screen-grid voltage of 250 volts.

As was stated before, Fig. 8 forms the first model of a relationship between cathode state and mutual conductance against which experimental results may be compared. Before such a comparison is made, however, it is proposed to derive a second model. The first is based on deterioration of the cathode due
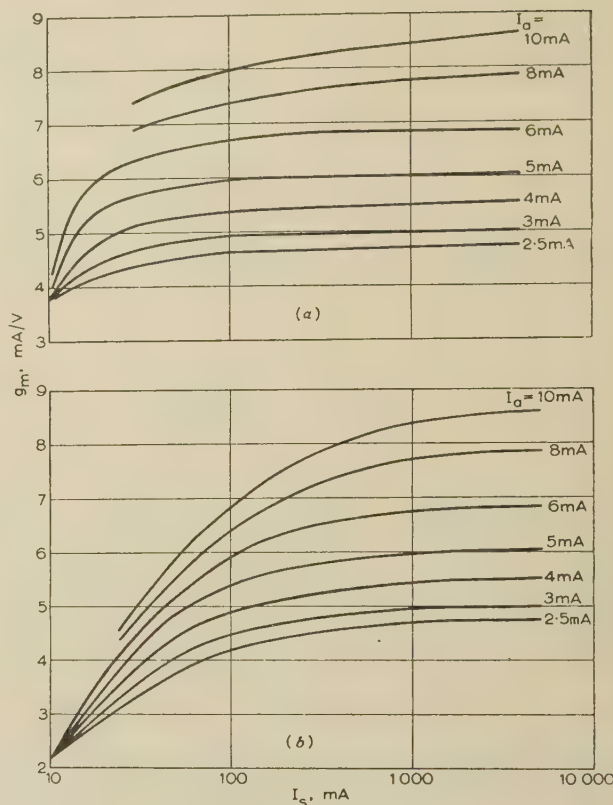
Fig. 9.—g_m/I_s characteristics, with I_a as parameter, for the 6P10 valve, forming the models linking the cathode state with the mutual conductance of the valve.

(a) First model.
(b) Second model.

to failing emission alone: the second will assume a decay in emission accompanied by an increase in cathode resistance. In order to form a reasonable estimate of the resistance to be associated with each level of emission in the first model, use is made of the experimental fact of inverse proportionality of emission and resistance derived from Reference 3 and from information provided by Hannay, MacNair and White,[10] although their interpretation of the underlying cause of inverse proportionality is not necessarily subscribed to. The cathode resistance for the second model is therefore calculated on the basis of the values of I_s used in the first model, in conjunction with eqn. (7). In order to formulate the second model it is possible to modify each I_a/V_{g1} characteristic in Fig. 8 by the addition of the appropriate cathode resistance, and then to produce from these a new set of g_m/I_s curves. Two such modified I_a/V_{g1} curves are included in Fig. 8. An alternative approach would be to modify Fig. 9(a) directly according to the equation

$$g_m = \frac{g'_m}{1 + R_K g'_m \alpha} \quad . \quad . \quad . \quad . \quad (12)$$

where g_m = New value of mutual conductance.
  g'_m = Value of mutual conductance in Fig. 9(a).
  R_K = Cathode resistance associated with the (g'_m, I_s) co-ordinates of Fig. 9(a), calculated according to eqn. (7).
  α = Ratio cathode-current/anode-current (~ 1.25).

Hence each co-ordinate (g'_m, I_s) of Fig. 9(a) would be translated to (g_m, I_s) in a new family of curves shown in Fig. 9(b), which

then forms a second model of a relationship between cathode state and mutual conductance. This model, however, takes account of reduction in emission together with a consequent increase in cathode resistance and is not restricted to the concept of falling emission alone as was the first model.
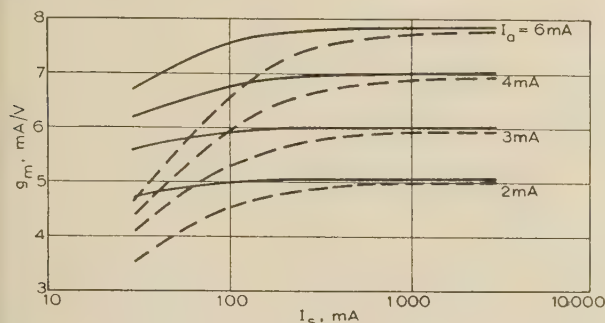
Fig. 10.—g_m/I_s characteristics, with I_a as parameter, for the 6P12 valve, forming the models linking cathode state with valve mutual conductance.

——————— First model.
— — — — Second model.

All the calculations necessary to provide these two models were then repeated for the 6P12. The two families of curves corresponding to Figs. 9(a) and 9(b) for the 6P10 are combined in Fig. 10 for the 6P12 for a screen-grid voltage of 60 volts.

(12) COMPARISON OF EXPERIMENTAL RESULTS WITH THE MODELS

Before any attempt at comparison is made between the models and experimental results taken from life-test records, one further difficulty must be resolved. This arises from the fact that the g_m/I_a relationship obtained from Langmuir's equation and valve design data, such as those suggested by Liebmann, is invariably more optimistic than the relationship obtained from measurement. This problem has been treated in unpublished work by Reynolds, and it is suggested that the discrepancy is due to difficulties in estimating the effective grid-cathode distance. Corrections due to variations in grid pitch, to cathode misalignment and to lack of control of the grid-wire diameter must also be considered in appreciating the difference between theory and practice.

For the purpose of this investigation it is proposed to overcome this difficulty by disregarding the parametric values of I_a in Figs. 9 and 10 in the following manner. The measured values of g_m, at constant I_a, and of the low-temperature total emission are extracted from the life-test records, and eqn. (3) is used to convert the low-temperature to the normal-temperature total emission. A plot of g_m against I_s at the several life-test stages can then be made using the same axes as Fig. 9 or 10. The shape of the resulting curve may be compared with the family of curves comprising the first model or that comprising the second. In particular, the resulting curve may be compared with one curve from the family of each model. The curve chosen in each case is that which passes through the highest g_m co-ordinate taken from the life-test records, neglecting the fact that the constant I_a value of the experimental curve will not necessarily coincide with the parametric value of I_a in the curves chosen from the models. It will be realized here that the curves chosen for comparison with the life-test plot may not be any of those actually drawn in Fig. 9 or 10. The curves shown are merely examples of a complete family whose individual separations may be made as small as required.

This decision to disregard the parametric value of I_a in the models can be justified by comparing the curve for 8 mA in Fig. 9(b) with the second model curve for 6 mA in Fig. 10. Here we have two different geometrical structures giving similar values of g_m for different values of I_a. Nevertheless, the two curves almost coincide, and therefore it can be said that the decay of g_m with I_s, from initial values of 7·8 mA/volt and 1 amp, is independent of whether the parametric I_a value is 8 or 6 mA. That this can be said of two differently designed structures would make it even more applicable when considering minor modifications to the same structure.

The comparison between models and life-test records is first made in Fig. 11(a). Here the average values of g_m and I_s at
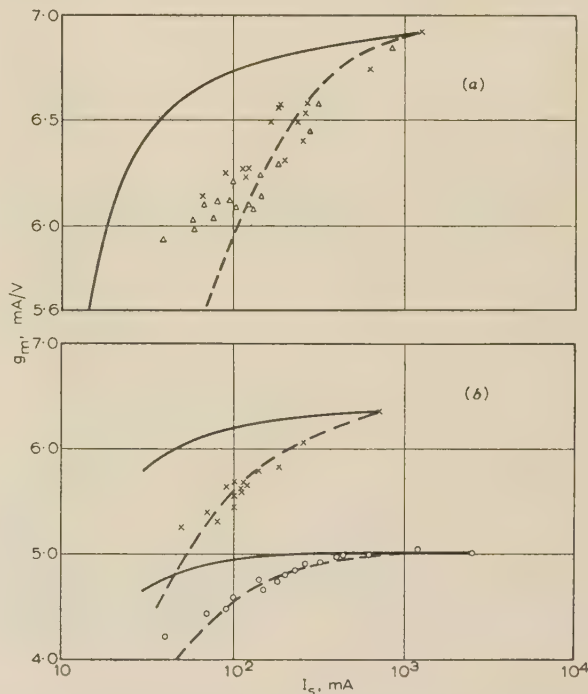
Fig. 11.—Comparison of two sets of average life-test measurements of g_m and I_s with selected curves from the first and second models.

——————— First model.
— — — — Second model.
(a) 6P10 valves.
(b) 6P12 valves.

several different stages of life test are obtained from the records of two batches of 10 valves of type 6P10. After eqn. (8) has been used to obtain an estimate of I_s at normal temperatures, the separate (g_m, I_s) co-ordinates for the two batches are plotted as crosses and triangles and compared with two model curves chosen in the manner described. In view of the fact that the highest (g_m, I_s) co-ordinates for each batch are almost coincident, a single curve from each model is shown for comparison. The life-test conditions in both cases were $V_a = V_{g2} = 250$ volts, $V_h = 6·3$ volts and $I_a = 8$ mA.

A second comparison is made in Fig. 11(b) for two batches of valves of type 6P12, one of 50 valves, plotted as crosses, and one of 10 valves, plotted as circles. In this case a separate pair of model curves is required for each batch. The life-test conditions were $V_a = 90$ volts, $V_{g2} = 60$ volts, $V_h = 5·5$ volts and $I_a = 6$ mA for the larger batch, and $V_a = V_{g2} = 40$ volts, $V_h = 5·5$ volts and $I_a = 3$ mA for the smaller batch. At this stage two points should receive comment: first, a heater voltage of 6·3 volts for the 6P10 would give a cathode-surface temperature of about 1 020° K, which is the value chosen for the

models (a heater voltage of 5·5 volts for the 6P12 would give a cathode-surface temperature of about $1010°$ K, or $10°$ lower); second, the screen-grid voltage used in the life test of the smaller batch of 6P12 valves was 40 volts whereas that used for the models was 60 volts. Neither the difference in temperature nor the difference in voltage is thought to be significant: the temperature difference is considered to be negligibly small, and the voltage difference can be shown to be unimportant through examination of the g_m/I_a characteristics of valves with V_{g2} as parameter. These curves show that, although the value of g_m for any particular value of I_a differs with V_{g2}, the rate of change of g_m with I_a at particular values of I_a is substantially independent of V_{g2} in the acceptable operating region of the valve. This means that a change in V_{g2} merely produces a slightly different basic I_a/V_{g1} shape, equivalent to a minor geometrical modification. Thus, by disregarding the parametric values of I_a in the models, it is possible to disregard, not only geometrical uncertainties in the valve structure, but also differences in screen-grid voltage.

Although the experimental points in Fig. 11 are derived from life-test records, no reference to time is made in either Figure. In order to present the results in another way, the plots in Fig. 11(b) are represented as life-test curves in Fig. 12, where the

obtained are fairly closely in line with the second model. In the following extension to much larger numbers of valves, however, modifications to eqn. (8) will be considered.

It is now proposed to consider life-test results on a much larger scale. The records of about 1 000 valves are examined over life-test periods ranging from 2 000 to 23 000 hours. These valves (and the preceding 80) were constructed to test the pre-processing of valve components and the use of non-standard piece-parts. They are all particularly useful for the present investigation, since the falls in g_m and I_s are, in general, considerably greater than those in standard valves of the same type. They therefore give greater opportunity for comparison between model and life-test records over substantially wider emission ranges.

The following method is used in the larger-scale investigation. First, the results presented in Figs. 11 and 12 are deemed sufficient to exclude the first model in favour of the second. This exclusion is supported in the review described below, where it can be shown that work functions less than 1 eV would be necessary to fit the life-test results to the first model. Such work functions are outside the range of those accepted in the literature or demonstrated in Part 1. (This point will be expanded in the next Section.) Consequently, in what follows, it will be
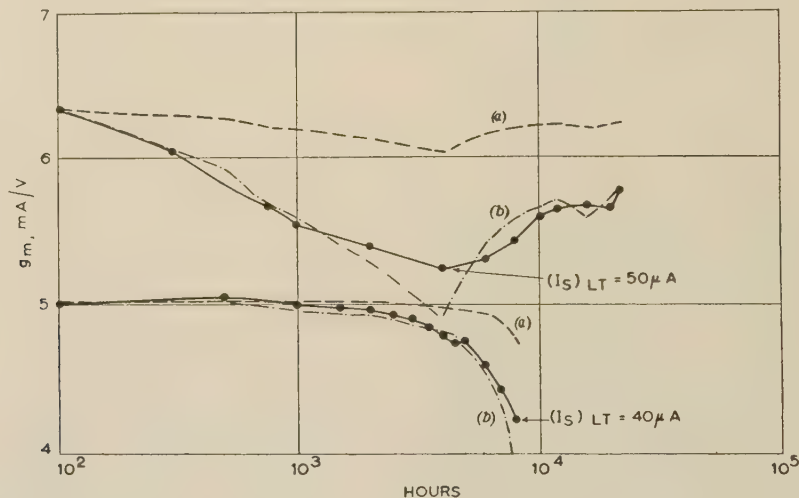
Fig. 12.—Comparison of experimental and calculated life characteristics of g_m for two sets of 6P12 valves.

The curves marked (a) indicate a calculation based on the application of the first model to life-test measurements of emission; the curves marked (b) indicate a similar calculation using the second model.

——————— Experimental values.
– – – – –
– · — · — } Calculated values.

full curves are the actual measurements of g_m taken during life test; the broken curves are the values of g_m computed from the life-test low-temperature emission results using eqn. (8) in conjunction with Langmuir's equation and Liebmann's design data, i.e. using the first model; and the dot-dash curves give g_m computed in the same way, through use of the second model. The computation is simply done by choosing the model curve which passes through the first (g_m, I_s) co-ordinate of the life-test record and subsequently reading off the g_m values for the sequence of emission levels of the life test.

In order to make the comparisons described in the preceding paragraphs, the actual measurements of low-temperature total emission are modified by eqn. (8) to translate them to normal temperatures before any comparison with a model is possible. For the small sample of 80 valves so far considered it would appear that use of this equation is not unreasonable when the low-temperature emission exceeds $150\,\mu A$, since the results

assumed that only the second model, incorporating the idea of cathode resistance, need be considered. The life-test results are then taken in batches of 10–100 valves. The average values of g_m and low-temperature total emission at the start of life are used in conjunction with an equation of the form

$$(I_s)_N = q(I_s)_{LT} \quad . \quad . \quad . \quad . \quad . \quad (13)$$

similar to eqn. (8) to locate a particular model curve in Figs. 9 or 10. The choice of a specific value for q is deferred for a moment, but, assuming a choice is made, the average value of the end-of-life emission is also used in eqn. (13), with the same value of q, to determine an estimate of the end-of-life normal-temperature emission. This emission is used to determine the final g_m on the particular curve chosen, i.e. according to the second model, and this g_m is compared with the final g_m taken from the life-test records according to the equation

Percentage agreement between life test and model =

$$100\left[1 - \frac{|\text{Final } g_m \text{ from life test} - \text{Final } g_m \text{ from model}|}{\text{Fall in } g_m \text{ from life test}}\right]$$

. . . . . . (14)

No account is taken of whether the second term inside the bracket has a positive or negative sign.

Before any results are presented, it is proposed to refer back to the choice of q in eqn. (13). When this larger-scale investigation based on the life-test records began it was found that, although a figure of 1 000 [cf. eqn. (8)] could be used for q very satisfactorily for a large number of 6P10 valves, a different value was more successful for other 6P10 valves and for 6P12 valves. It was therefore decided to divide the batches into 10 groups, according to valve types, life-test conditions, quality and final value of low-temperature emission. For each of the groups a value of q was chosen to the nearest 100 to bring the agreement expressed in eqn. (14) to at least about 90%. When making this choice it will be appreciated that the process of locating the model curve gives little information on the necessary value of q which must be adopted. This is particularly true when the initial low-temperature emission is greater than 1 mA. For example, the curve located by a g_m of 6·5 mA/volt and a low-temperature emission of 1 mA associated with a q of 500 is not very far removed from the curve located by the same g_m and emission associated with a q of 1 500. The reason for this is, of course, the fact that the model curves are all nearly parallel to the emission axis for emissions greater than 1 amp. As the emission diminishes and the model curves tend to become parallel with the g_m axis, so the choice of a value for q becomes more decisive in locating the required model curve.

Clearly q must finally be chosen to bring the recorded g_m at the end of life so close to a particular model curve that the 90% agreement is achieved. The particular model curve is located in the first place by the initial values of g_m and low-temperature emission in conjunction with eqn. (13) and any arbitrary value of q in the right range. A figure of 1 000 is quite reasonable here. In Table 1, after a final choice of q has been made, a minor

records and the second model. It is therefore not possible to justify the second model on the basis of agreement with experiment results, except in so far as it can be admitted that agreement is possible if the appropriate adjustments are made to q. The validity of the second model must stand or fall by whether the q modifications can be justified experimentally or on accepted previous experience of the underlying principles of the cathode. This aspect is treated in the next Section.

It should be emphasized that no special attempt was made to adjust q more precisely than to the nearest 100, although a closer percentage agreement could have been obtained in this way. The value of $\phi^{(0)}$ was calculated according to the equation

$$\log_\varepsilon q = \frac{e\phi^{(0)}}{k}\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \quad . \quad . \quad . \quad . \quad (15)$$

derived from eqn. (6) and reducing to

$$\log_\varepsilon q = 11\cdot6\phi^{(0)} \times 0\cdot45 = 5\cdot22\phi^{(0)}$$

with $T_1 = 700°\text{K}$ and $T_2 = 1\,020°\text{K}$.

Having presented these results, it must also be recorded that in addition there was a small group of valves, about 30 in all, which could be fitted to the second model only if a very low value of q were assumed. These valves had been given an unsuitable heater and their life-test records showed normal emission characteristics with very serious reductions in g_m (by about 20%) in the first 1 000–2 000 hours' life. The low value of q involved a value of $\phi^{(0)}$ on the lower fringe of, or entirely outside, the distribution of $\phi^{(0)}$ given in Part 1.

A small percentage of the valves (less than 5%) with results presented in Table 1 had increasing values of low-temperature emission for the first few hundred hours of life. During these periods the mutual conductance showed either no change or a slight fall. These periods have been regarded as transients, and the initial values of emission and mutual conductance have been taken at the peak of the emission/time characteristic.

Finally, a short life test was conducted on five valves in which the value of $\phi^{(0)}$ was measured. It was shown here that in the

Table 1

Valve type and life-test conditions	Group				Number of valves	Agreement [eqn. (14)]	q	$\phi^{(0)}$
6P10 $V_a = 250$ volts $V_{g2} = 250$ volts $I_a = 6$ mA $V_h = 6\cdot3$ volts	(1) Final $(I_s)_{LT} > 100\,\mu\text{A}$ ..	..	..	..	130	% 87	1 000	1·32
	(2) Final $(I_s)_{LT} = 80–100\,\mu\text{A}$ ..	..	..	..	20	97	1 500	1·40
	(3) Final $(I_s)_{LT} = 35\,\mu\text{A}$ ..	..	..	..	10	98	3 500	1·56
6P12 $V_a = 40$ volts $V_{g2} = 40$ or 60 volts $I_a = 3$ or 4 mA $V_h = 5\cdot5$ volts	(4) Final $(I_s)_{LT} > 150\,\mu\text{A}$ ..	..	..	..	380	87	500	1·19
	(5) Final $(I_s)_{LT} = 100–150\,\mu\text{A}$ ..	..	..	..	30	86	1 000	1·32
	(6) Final $(I_s)_{LT} = 30–50\,\mu\text{A}$ ..	..	..	..	30	85	1 500	1·40
6P12 $V_a = 90$ volts $V_{g2} = 60$ volts $I_a = 6$ mA $V_h = 5\cdot5$ volts	(7) Final $(I_s)_{LT} > 150\,\mu\text{A}$ (Later samples) ..	..			160	88	300	1·09
	(8) Final $(I_s)_{LT} > 150\,\mu\text{A}$ (Early samples A)	..			100	83	400	1·15
	(9) Final $(I_s)_{LT} > 150\,\mu\text{A}$ (Early samples B)..				70	87	500	1·19
	(10) Final $(I_s)_{LT} < 150\,\mu\text{A}$	..	..		50	98	800	1·28

correction is made to the initial choice of model curve by substituting the final value q for the arbitrary choice in the context of the initial g_m and emission values.

It will be observed from the previous two paragraphs that agreement to within 90% has been forced between life-test

initial stages of life a 50% fall in emission in the region 1–5 mA was due more to the movement of the log $I/(1/T)$ characteristic (RR'R''R''' in Fig. 5) to the left, keeping a substantially constant slope, than to any change of slope or work function. The average value of $\phi^{(0)}$ recorded was 1·15 ± 0·03 volts at each stage.

(13) DISCUSSION

The bulk of the experimental results given in the previous Section tend to show that the measurements of mutual conductance and low-temperature total emission recorded during life can be aligned with a model of a relationship between the cathode state and the mutual conductance of the valve based on the following four assumptions:

(a) That the Langmuir relation for a planar diode can be applied to the equivalent-diode concept of a normal pentode receiving valve.

(b) That pentode design data, e.g. those given by Liebmann, give a reasonably close approximation to the valve characteristics, and that the final necessary approximation can be made by assuming that the initial life test values of g_m and normal-temperature total emission locate the model curve, irrespective of non-coincidence of I_a values.

(c) That the normal-temperature total emission may be derived from the low-temperature measurement through use of an equation such as eqn. (13).

(d) That a resistance is associated with the cathode which has a value given by eqn. (7).

It is considered that assumptions (a), (b) and (d) rest on a much firmer foundation than does (c) and that, although the model can be aligned with the experimental results on the basis of the four assumptions taken together, the validity of the alignment must rest on a closer inspection of the assumption (c).

Before taking the third assumption in more detail, a comment on the fourth is necessary. The qualitative nature of the interdependence of cathode emission and resistance has a greater weight of supporting evidence than has the quantitative value of the constant of proportionality given in eqn. (7). The figure of 1·6 arises, as explained before, from the work done in Reference 3, and some of the weaknesses of the method have been pointed out. A little extra support comes from the fact that for an emission of 1 amp/cm² —a figure accepted as representative for oxide-coated-cathodes at 1 000° K—eqn. (7) gives a cathode resistance of 1·6 ohms for a cathode of 1 cm.² This in its turn is acceptable, giving a conductivity of about 4×10^{-3} mho/cm. A still further point to be taken into account is that an increase or decrease in the factor by 50%, i.e. a change to 2·4 or 0·8, would cause the change of q needed to bring the life-test records into line with the model. Such a change in q would involve a change in $\phi^{(0)}$ of about ±0·05 eV at an emission of about 100 μA. These changes could bring the value of $\phi^{(0)}$ for group (4) in Table 1 very close to 1·0 eV and for group (2) to greater than 1·6 eV. As will be shown later, such values are on the limits of acceptable work functions. Any larger changes in the constant in eqn. (7) would involve unacceptable values of $\phi^{(0)}$ if agreement with the second model were required.

Coming now to the third assumption and taking the experimental results in sequence, the graphical comparisons in Figs. 11 and 12 show clearly that, if a q value of 1 000 is used, a model based on the four assumptions gives a better alignment with the life-test records than does a model based on the first three assumptions, at least for emissions less than, say, 750 mA. A q factor of 1 000, giving $\phi^{(0)} = 1 \cdot 32$ eV, corresponds to the centre of the distribution in Part 1 and is thus acceptable as a reasonable assumption. A consequent $\phi^{(5/4)}$ value of 1·2 eV is well within the range of values for the (BaSr)O cathode quoted in recent literature. Herrmann and Wagener[11] give a Table of values of $\phi^{(2)}$ ranging from 1·0 to 1·65 eV obtained by different investigators over the past 30 years. As was mentioned briefly before, in order to force the life-test records into alignment with the first model, very low q values, involving $\phi^{(0)}$ values lower than 1 eV, would be required. These are not obtained in the experimental work in Part 1, nor in the results tabulated by Herrmann and Wagener. Consequently, the first model is rejected for the range of life-test records presented here.

Figs. 11 and 12 also show that the second model apparently breaks down when the low-temperature total emission falls below about 150 μA. It is, however, possible to re-align records and model if it is assumed that the q factor or work function increases when the emission falls below such a level. This concept will be supported during discussion of the Table of life-test results. An increase of work function with reduction of emission is acceptable, being consistent with the general appreciation of the nature of emission from both metals and semi-conductors. Nevertheless it should be noted that the increase of work function, although coinciding with decreasing emission at about 700° K, actually counterbalances the decrease when referring to 1 000° K. This has been noted in extreme cases when running pentodes continuously with anodes at red heat (6 watts in place of the normal 1·5 watts). Here, despite a level of low-temperature emission of about 3 μA, almost completely normal measurements of mutual conductance were made at 1 000° K for the whole period of the test (about 5 000 hours). Without going into detail, these results could be due to balancing changes in the $A^{(n)}$ and $\phi^{(n)}$ factors in eqn. (6), and the present method may provide a way of studying such changes, making use of the large samples available in life-test records.

It must be emphasized that the fact that a constant value of q gives reasonable alignment between model and records for emissions greater than 200 μA cannot stand alone in justifying the hypothesis of a constant work function in this region. It has been pointed out before that, for the higher emission levels, the value of q is not very significant in locating model curves. Hence the alignment of model and records in this region cannot be used to support constancy of q factor, even though a constant q has in fact been used. This is, of course, due to the smoothing effect of the space-change cloud, and is just another way of saying that, for high emission levels, the mutual conductance is largely independent of emission. At the same time it must be remembered that the short life test, where $\phi^{(0)}$ values were measured over a period of time, indicates that there is a possibility of emission changes occurring at these high levels without consequent change in q or $\phi^{(0)}$.

Table 1 will now be considered. The weighted average value of $\phi^{(0)}$ for all the valves is 1·21 eV, and this gives an average agreement of 88% between model and life-test records. The average $\phi^{(0)}$ value for the main part of the distribution in Fig. 7 is 1·35 eV, which corresponds reasonably well considering the increase of 36 times in the size of the sample. The distribution of average $\phi^{(0)}$ values for the 10 groups in the Table ranges from 1·09 to 1·56 eV, and this again is in reasonable agreement with the main distribution and with the Table of values given by Herrmann and Wagener.

Considering the 10 groups in three sets (1)–(3), (4)–(6) and (7)–(10), it will be seen that as the emission falls there is evidence in each set that higher values of q, and therefore of $\phi^{(0)}$, are required to align the model and the records. Remembering again that little significance can be attached to the use of the same q factor for the initial and final results within each group, the results of the three sets give evidence of increasing work functions as soon as the lower emission levels are reached. This is in agreement with the conclusions reached when considering the graphical comparisons in Figs. 11 and 12, and also with the rather slight evidence of increasing work function with life obtained in Part 1.

It will also be observed that for the 6P12 valves the average $\phi^{(0)}$ value is 1·19 eV for group (4) and 1·13 eV for groups (7), (8) and (9), comparing groups with similar final low-temperature emissions but different life-test conditions. In the same way the average $\phi^{(0)}$ value is 1·36 eV for groups (5) and (6) and 1·28 eV for group (10). It would therefore appear that for similar

emission levels the work functions necessary to align groups (4)–(6) with the second model are higher than those required to align groups (7)–(10). This is interesting when taken in conjunction with life-test evidence not presented in Section 12. It has been found when life testing higher-quality valves for submarine repeaters that input valves, life tested under the conditions used for groups (4)–(6), always settle down after about 5000 hours to a low-temperature emission level which is about 70% of that achieved by output valves life tested under the conditions used for groups (7)–(10). The work functions $1 \cdot 19$ and $1 \cdot 13$ eV correspond with q values of 500 and 360. Taking 700 μA as a final figure for the low-temperature emission of higher-quality output valves and 490 μA for similar input valves, these q factors would give estimates of normal-temperature emission of 250 mA for output valves and 245 mA for input valves. Without attaching too much importance to the actual figures, these estimates are more reasonable than the low-temperature levels, since a substantial case can be made to show that the input conditions are less onerous to the cathode state than the output conditions.

There remains to consider that group of valves with unsuitable heaters which could be fitted to the second model only if a very low value of q were assumed. This group bears the same relation to the second model as do the rest of the valves to the first. The medium deteriorations of g_m for the bulk of the valves need very low values of q to produce agreement with the first model. The abnormally high deteriorations of g_m for the group under consideration need, at low emission levels, very low values of q (about 100) to produce agreement with the second model, and absurdly low ones (about 20) for agreement with the first model. The corresponding $\phi^{(0)}$ values are $0 \cdot 88$ and $0 \cdot 57$, both of which are rejected as being unrealistic. These valves also require the peculiar coincidence of a reduction in $\phi^{(0)}$ as the emission decays. Thus, although agreement can be forced, it is preferred to argue that there is a form of deterioration of cathode state which is not in accord with any model proposed here. It could also be argued that failure to fit the second model coincides with a dangerously unsatisfactory cathode state, although successful alignment with the second model does not imply a satisfactory cathode. The behaviour of this group may be contrasted with group (10) in Table 1, whose life-test curve is plotted in Fig. 12 (upper curve). It will be observed that the initial deteriorations in g_m are very comparable, but that the behaviour on subsequent life test is very different. The valves in group (10) are showing signs of complete recovery after more than 20 000 hours, and are also in accord with the second model at all stages of life; the valves in the group under discussion here fail completely before 10 000 hours, and are at no stage in reasonable accord with any model presented here.

Although it is not proposed to suggest any further model at this stage to cover the exceptions in the previous paragraph, it may be useful to indicate one direction in which the solution may be. Each of the four assumptions on which the second model is based tacitly implies a homogeneous cathode state. It may be that the cause of the deviations lies in non-uniformity of some part of the cathode structure in the presence of particular forms of severe cathode deterioration. This suggestion must, however, await further experimentation before any confidence can be placed in it.

Before leaving this discussion it must be emphasized that all the conclusions reached relate to emission levels lower than about 750 mA. The technique of investigating the relationships between the cathode state and the mutual conductance of the valve, in order to construct a model for the linkage between them, cannot be applied for emissions higher than this level. In such a region changes in emission or in cathode resistance, if

inversely proportional to emission, are insufficient to affect the mutual conductance which is used as a yardstick for the whole investigation. Indeed, it will be observed that the first and second models are almost indistinguishable at emission levels greater than 1 amp.

(14) CONCLUSIONS

A method of deriving the normal-temperature zero-field emission from the oxide cathodes of receiving valves was described in Part 1. The method made use of extrapolation from the low-temperature emission measurement and yielded reasonable values of the work function in the process.

The problems of measurement of the cathode resistance of normal receiving valves are by no means solved. It is, however, possible to derive a relationship, based on experiment, between emission and resistance which enables an estimate of the latter to be made in terms of measurements of the former; this necessitates a link between low-temperature and normal-temperature total emission from the cathode, which is established empirically from the work done in Part 1, and introduces the concept of cathode work-function.

The combined effect of cathode emission and cathode resistance can then be used to form a model of the relationship between the cathode state and the 4-terminal mutual conductance of the valve. The model takes account of the current/voltage relationship of a planar diode developed by Langmuir and of valve design data such as those suggested by Liebmann.

The validity of the model is checked first by showing that the life-test records can be manipulated in such a way as to conform with the model, but the real test lies in examining the manipulations which have to be made to produce identity between the model and the records. These manipulations can be reduced to assumptions relating to the cathode work-function, but it can be shown that in nearly all cases these assumptions can be justified through reference to Part 1, or to values of work function accepted in the literature, or to the underlying principles of emission from metals and semi-conductors.

In a limited number of cases the model fails to account for the life-test records because the necessary manipulations of the cathode work-function are unacceptable. The failure may lie in the inhomogeneity of the cathode under certain conditions of severe deterioration of cathode state, but more experimental evidence is required to prove this.

It is hoped that the methods developed may prove useful in permitting the large bulk of evidence amassed in the form of life-test records to be used in attempts to understand the nature of the oxide cathode.

(15) ACKNOWLEDGMENTS

(16) REFERENCES

(1) EISENSTEIN, A.: 'The Leaky-Condenser Oxide Cathode Interface', *Journal of Applied Physics*, 1951, **22**, p. 138.
(2) LOOSJES, R., and VINK, H. J.: 'The Conduction Mechanism of Oxide-Coated Cathodes', *Philips Research Reports*, 1949, **4**, p. 449.
(3) METSON, G. H., WAGENER, S., HOLMES, M. F., and CHILD, M. R.: 'Life of Oxide-Coated Cathodes in Modern Receiving Valves', *Proceedings I.E.E.*, Paper No. 1137 R, June, 1951 (**99**, Part III, p. 69).
(4) DALMAN, G. C.: 'Effects of Cathode and Anode Resistance on the Retarding Potential Characteristics of Diodes', *Journal of Applied Physics*, 1954, **25**, p. 1263.

(5) NERGAARD, L. S.: 'Studies of the Oxide-Cathode', *RCA Review*, 1952, **13**, p. 464.

(6) TILLMAN, J. R., BUTTERWORTH, J., and WARREN, R. E.: 'The Dependence of Mutual Conductance on Frequency of Aged Oxide-Cathode Valves and its Influence on their Transient Response', *Proceedings I.E.E.*, Monograph No. 56, December, 1952 (**100**, Part IV, p. 8).

(7) LIEBMANN, G.: 'The Calculation of Amplifier Valve Characteristics', *Journal I.E.E.*, 1946, **93**, Part III, p. 138.

(8) GYSAE, B., and WAGENER, S.: *Zeitschrift für Technische Physik*, 1938, **19**, p. 264.

(9) FERRIS, W. R.: 'Some Characteristics of Diodes with Oxide-Coated Cathodes', *RCA Review*, 1949, **10**, p. 134.

(10) HANNAY, N. B., MacNAIR, D., and WHITE, A. H.: 'Semi-Conducting Properties in Oxide Cathodes', *Journal of Applied Physics*, 1949, **20**, p. 669.

(11) HERRMANN, G., and WAGENER, S.: 'The Oxide-Coated Cathode' (Chapman and Hall, 1951), Volume II, p. 181.

TRANSFORMER-TYPE EQUIVALENT NETWORKS

By Professor C. E. MOORHOUSE, B.Mech.E., M.E.E., Member.

SUMMARY

Various equivalent circuits used, or suggested, for transformers and induction motors are examined by considering them as 4-terminal networks, and determining the corresponding general circuit constants by matrix methods. The information obtainable from tests is discussed, and the use of the 4-terminal network is proposed as considerable generality is achieved by this approach.

LIST OF SYMBOLS

Input Quantities.

V_1 = Input voltage.
I_1 = Input current.
Z_{IN} = Input impedance.
Y_{IN} = Input admittance.

$[S_1] = \begin{bmatrix} V_1 \\ I_1 \end{bmatrix}$ = Input quantity matrix.

Output Quantities.

V_2 = Output voltage.
I_2 = Output current.
Z_L = Terminating impedance.

$[S_2] = \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}$ = Output quantity matrix.

Other Symbols.

A, B, C, D = General circuit constants.

$[O] = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ = General circuit constant matrix.

k = Complex transformation ratio.
k = Turns ratio.

(1) INTRODUCTION

The transformer and the induction motor have been represented by equivalent circuits of various kinds, all composed of linear circuit elements.[3,4,8] Some of these are referred to as 'exact', some as 'practical', some as 'approximate'.

As these equivalent circuits are all passive networks of the ladder type, consisting of successive shunt and series components, it is logical to apply matrix methods to their analysis considering them as 4-terminal networks.[1]

(2) THE FOUR-TERMINAL NETWORK APPROACH

Any linear electrical device having an input and an output can be represented by a 4-terminal network (Fig. 1) whose behaviour is described by the equations

$$V_1 = AV_2 + BI_2 \quad \ldots \quad \ldots \quad (1)$$

$$I_1 = CV_2 + DI_2 \quad \ldots \quad \ldots \quad (2)$$
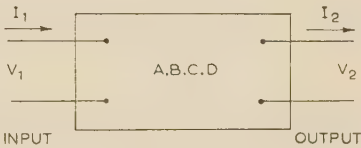
Fig. 1

where A, B, C and D are called the 'general circuit constants' and may be regarded as complex operators like Z and Y.

If the device is a passive one the further relation

$$AD - BC = 1 \quad \ldots \quad \ldots \quad \ldots \quad (3)$$

is found to hold, so that only three of the quantities are independent in this case.

Also

$$V_2 = DV_1 - BI_1 \quad \ldots \quad \ldots \quad \ldots \quad (4)$$

$$I_2 = -CV_1 + AI_1 \quad \ldots \quad \ldots \quad \ldots \quad (5)$$

If the device is symmetrical, i.e. if interchanging the input and output terminals makes no difference, $A = D$ and only two of the quantities are independent.

The nature of the general circuit constants can be well illustrated by an examination of some particular cases.

(*a*) If $I_2 = 0$, $A = V_1/V_2$ and $C = I_1/V_2$ [from eqns. (1) and (2)] so that A is the open-circuit (complex) voltage ratio and C is an open-circuit transfer admittance.

(*b*) If $V_2 = 0$, $B = V_1/I_2$ and $D = I_1/I_2$ [from eqns. (1) and (2)] so that B is a short-circuit transfer impedance and D is the short-circuit (complex) current ratio.

The general circuit constants can thus be determined from open-circuit and short-circuit tests on the device concerned, provided that magnitudes and phase angles are measured.

They can also be found by manipulation of the equations of performance of the device, and by computation by matrix methods from the equivalent circuit.

Any given passive linear 4-terminal network, i.e. with given values of three of A, B, C and D, can be represented by a large number of equivalent circuits; impedances, admittances and complex ratio changers being employed. Of these equivalent circuits the simplest will contain three complex elements—two configurations, the T and π, being possible in any case, although not necessarily both physically realizable.

For any 4-terminal network the general circuit constant matrix is defined by

$$[O] = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

(2.1) The Coupled Circuit

As an illustration of the method of finding the general circuit constants from the equations of performance consider the coupled circuits shown in Fig. 2. Here the usual equations are:

$$V_1 = Z_{11}I_1 + Z_{12}I_2 \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$-V_2 = Z_{21}I_1 + Z_{22}I_2 \quad . \quad . \quad . \quad . \quad . \quad (7)$$

$$(\text{or } 0 = Z_{21}I_1 + Z_{22}I_2 + V_2)$$

where $Z_{11} = r_1 + j\omega L_1$, $Z_{22} = r_2 + j\omega L_2$, $Z_{12} = Z_{21} = j\omega M$
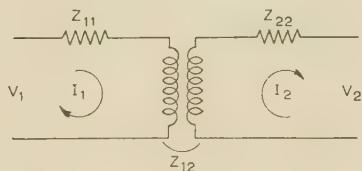
Fig. 2

When $\qquad I_2 = 0$

$$A = \frac{V_1}{V_2} = -\frac{Z_{11}}{Z_{12}}$$

and

$$C = \frac{I_1}{V_2} = -\frac{1}{Z_{12}}$$

When $\qquad V_2 = 0$

$$D = \frac{I_1}{I_2} = -\frac{Z_{22}}{Z_{12}} \text{ (since } Z_{21} = Z_{12})$$

also

$$B = \frac{AD - 1}{C} = \frac{-Z_{11}Z_{22} + Z_{12}^2}{Z_{12}}$$

(2.2) Some Simple Circuits

For the simple (degenerate) 4-terminal network the values of the general circuit constants may be written down from inspection.[2]

Series Impedance.

$$\left.\begin{array}{c} V_1 = V_2 + ZI_2 \\ I_1 = \qquad I_2 \end{array}\right\} [O] = \begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$$

Shunt Admittance.

$$\left.\begin{array}{c} V_1 = V_2 \\ I_1 = YV_2 + I_2 \end{array}\right\} [O] = \begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$$

Ideal Ratio Changer.

(*k* may be either real or complex.)

$$\left.\begin{array}{c} V_1 = kV_2 \\ I_1 = \dfrac{I_2}{k} \end{array}\right\} [O] = \begin{bmatrix} k & 0 \\ 0 & \dfrac{1}{k} \end{bmatrix}$$

(2.3) Computation for More Elaborate 4-Terminal Networks of the Ladder Type[1]

This is most conveniently performed by successive matrix multiplication, the original eqns. (1) and (2) being replaced by the matrix equation

$$[S_1] = [O][S_2] \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where $\quad [S_1]$ is the column matrix $\begin{bmatrix} V_1 \\ I_1 \end{bmatrix}$

$\qquad [S_2]$ is the column matrix $\begin{bmatrix} V_2 \\ I_2 \end{bmatrix}$

and $\qquad [O]$ the general circuit constant matrix.

For two 4-terminal networks a, b in tandem

$$[S_{1a}] = [O_a][S_{2a}]$$

$$[S_{1b}] = [O_b][S_{2b}]$$

Also $[S_{2a}] = [S_{1b}]$ and hence $[S_{1a}] = [O_a][O_b][S_{2b}] \quad . \quad (9)$

$$[O] \text{ equivalent} = [O_a][O_b]$$
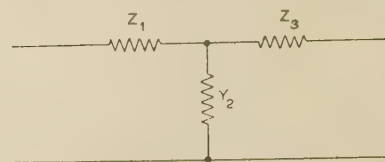
Fig. 3

The T-Network (Fig. 3).

$$[O_T] = \begin{bmatrix} 1 & Z_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Y_2 & 1 \end{bmatrix} \begin{bmatrix} 1 & Z_3 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 + Z_1 Y_2 & Z_1 + Z_1 Y_2 Z_3 + Z_3 \\ Y_1 & 1 + Y_2 Z_3 \end{bmatrix}$$

If three of A, B, C and D are given, appropriate values can immediately be assigned to Z_1, Y_2, and Z_3, i.e. the corresponding T-network can be designed

$$\begin{bmatrix} 1 + Z_1 Y_2 & Z_1 + Z_1 Y_2 Z_3 + Z_3 \\ Y_2 & 1 + Y_2 Z_3 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

from which $Y_2 = C$

$$Z_1 = \frac{A - 1}{C}$$

$$Z_3 = \frac{D - 1}{C}$$

Fig. 4

The π-Network.
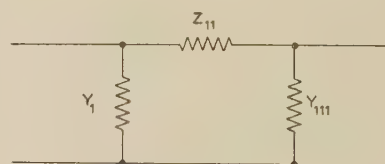
The π-network (Fig. 4) is treated similarly:

$$[O_\pi] = \begin{bmatrix} 1 + Z_{11} Y_{111} & Z_{11} \\ Y_1 + Y_1 Z_{11} Y_{111} + Y_{111} & 1 + Y_1 Z_{11} \end{bmatrix}$$

from which $\quad Y_1 = \dfrac{D - 1}{B}$

$$Z_{11} = B$$

$$Y_{111} = \frac{A - 1}{B}$$

The Coupled Circuit.

For the coupled circuit shown in Fig. 2 in which

$$[O] = \begin{bmatrix} -\dfrac{Z_{11}}{Z_{12}} & \dfrac{-Z_{11}Z_{22} + Z_{12}^2}{Z_{12}} \\ -\dfrac{1}{Z_{12}} & -\dfrac{Z_{22}}{Z_{12}} \end{bmatrix}$$

the equivalent T-network gives

$$Z_1 = Z_{11} + Z_{12}$$

$$Y_2 = -Y_{12}$$

$$Z_3 = Z_{22} + Z_{12}$$

and the π-network gives

$$Y_1 = (-Z_{22} - Z_{12})/(-Z_{11}Z_{22} + Z_{12}^2)$$

$$Z_{11} = (-Z_{11}Z_{22} + Z_{12}^2)/Z_{12}$$

$$Y_{111} = (-Z_{11} - Z_{12})/(-Z_{11}Z_{22} + Z_{12}^2)$$

the complexity of the second result indicating why the equivalent π-network is not favoured for general use.

(2.4) The Terminated 4-Terminal Network

A terminating impedance Z_L across the output of the circuit shown in Fig. 1 imposes the relation $V_2 = Z_L I_2$ so that eqns. (1) and (2) become

$$V_1 = AZ_L I_2 + BI_2 \quad . \quad . \quad . \quad . \quad (1a)$$

$$I_1 = CZ_L I_2 + DI_2 \quad . \quad . \quad . \quad . \quad (2a)$$

and the input impdance becomes

$$Z_{IN} = \frac{V_1}{I_1} = \frac{AZ_L + B}{CZ_L + D} \quad . \quad . \quad . \quad (10)$$

The general circuit constants have been widely used in power system analysis for transmission problems. Tables of their values for transmission systems may be found in various handbooks,[3] but the rules for the combination of components are not obvious unless matrix methods are employed.

(3) THE POWER TRANSFORMER

(3.1) 'Exact' Equivalent Circuit

The 'exact' equivalent circuit of the power transformer, given in Fig. 5, contains three complex elements and a simple ratio changer (one more element than is needed).
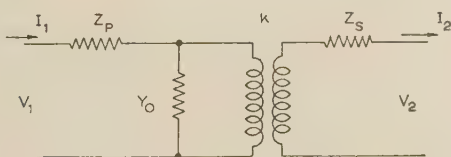
Fig. 5

For this, writing down the general circuit constants for the simple 4-terminal networks indicated,

$$[O] = \begin{bmatrix} 1 & Z_P \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Y_0 & 1 \end{bmatrix} \begin{bmatrix} k & 0 \\ 0 & \frac{1}{k} \end{bmatrix} \begin{bmatrix} 1 & Z_s \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} k(1 + Z_P Y_0) & (1 + Z_P Y_0)kZ_s + \dfrac{Z_P}{k} \\[2ex] kY_0 & kZ_s Y_0 + \dfrac{1}{k} \end{bmatrix}$$

Since this circuit contains four components, three complex (Z_P, Y_0, Z_s) and one real (k), it cannot be obtained from the measured general circuit constants unless one quantity is fixed.

If k is fixed (which is done by the designer) and A, B, C, D are given, the equivalent circuit then gives

$$Y_0 = \frac{C}{k}$$

$$Z_P = \frac{A - k}{C}$$

$$Z_s = \frac{D - \dfrac{1}{k}}{C}$$

(3.2) Exact Equivalent Circuit with Referred Quantities

Now, since

$$\begin{bmatrix} k & 0 \\ 0 & \frac{1}{k} \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & k^2 B \\ \dfrac{C}{k^2} & D \end{bmatrix} \begin{bmatrix} k & 0 \\ 0 & \frac{1}{k} \end{bmatrix}$$

it is possible to rearrange the circuit by placing the ratio changer in any desired position provided that the impedances (and admittance) affected are modified accordingly.

Thus if the ratio changer is moved to the right (after Z_s),

$$[O] = \begin{bmatrix} 1 & Z_P \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Y_0 & 1 \end{bmatrix} \begin{bmatrix} 1 & k^2 Z_s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k & 0 \\ 0 & \frac{1}{k} \end{bmatrix}$$

By writing

$$[S_1] = [O]'[S_2]'$$

where

$$[S_2]' = \begin{bmatrix} V_2' \\ I_2' \end{bmatrix} = \begin{bmatrix} k & 0 \\ 0 & \frac{1}{k} \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}$$

and

$$[O]' = \begin{bmatrix} 1 & Z_P \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Y_0 & 1 \end{bmatrix} \begin{bmatrix} 1 & k^2 Z_s \\ 0 & 1 \end{bmatrix}$$

the ratio changer is removed from further consideration and a simple 3-element circuit can be employed, the T-network being the most convenient. In this the circuit elements are now Z_P, Y_0 and $Z_s' = k^2 Z_s$, with the output current $I_2' = I_2/k$ and the output voltage $V_2' = kV_2$. This is the process usually called 'referring all quantities to the primary'. It involves the transfer of the ratio changer to the right and the use of transformed secondary voltage and current. (This is also what is achieved by the use of per cent or per unit quantities.)

This exact equivalent network is widely used; it is favoured as giving a physical picture of the transformer, and it is convenient because the quantities in it can be computed with reasonable accuracy by the designer.

It suffers from the drawbacks that after manufacture k can never be determined exactly (the quantity found by test being $|A|$) and that it is not possible to separate x_P and x_s the primary and secondary leakage reactances.

(3.3) The Approximate Equivalent Circuit

In the approximate equivalent circuit the exciting admittance Y is moved to the left (in front of Z):

$$[O]'_{APP} = \begin{bmatrix} 1 & 0 \\ Y_0 & 1 \end{bmatrix} \begin{bmatrix} 1 & Z_P \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & Z_s' \\ 0 & 1 \end{bmatrix}$$

In the case of the ordinary power transformer the errors produced by this process are small. Their effect, however, can readily be investigated by examination of the general circuit constants.

The exact value of $[O]'$ is $\begin{bmatrix} 1 + Z_P Y_0 & Z_P + Z_P Y_0 Z_s' + Z_s' \\ Y_0 & 1 + Z_s' Y_0 \end{bmatrix}$

and the approximate value

$[O]'_{APP}$ is $\begin{bmatrix} 1 & Z_P + Z_s' \\ Y_0 & 1 + Z_P Y_0 + Z_s' Y_0 \end{bmatrix}$

The differences between the exact and the approximate values are

$$\Delta A = Z_P Y_0, \quad \Delta B = Z_P Y_0 Z_s', \quad \Delta C = 0, \quad \Delta D = Z_P Y_0$$

and the proportional errors in magnitude are

$$\varepsilon_A = \left| \frac{\Delta A}{A} \right| = \left| \frac{Z_P Y_0}{1 + Z_P Y_0} \right|$$

$$\varepsilon_B = \left| \frac{\Delta B}{B} \right| = \left| \frac{Z_P Y_0 Z_s'}{Z_P + Z_P Y_0 Z_s' + Z_s'} \right|$$

$$\varepsilon_D = \left| \frac{\Delta D}{D} \right| = \left| \frac{Z_P Y_0}{1 + Z_s' Y_0} \right|$$

For a typical 10 MVA 3-phase 50 c/s 22/6·6 kV delta-star transformer the percentage values would be approximately

$$Z_P = Z_s' = 0·3 + j4·0$$
$$Y_0 = 0·0018 - j0·0108$$

Hence $Z_P Y_0 = 0·044 + j0·0041$ so that the errors in all three general circuit constants are small in this case.

For the power transformer, then, approximately,

$$\varepsilon_A = |Z_P Y_0|, \quad \varepsilon_B = \left| \frac{Z_P Y_0}{2 + Z_P Y_0} \right|, \quad \varepsilon_D = |Z_P Y_0|$$

The approximate 'referred' circuit reduces the T-network of the exact 'referred' circuit to an L or 'cantilever' with two components only and may hence be readily computed from test results.

(3.4) Other Equivalent Circuits

The common 'exact' equivalent circuit is only one of many possible 'exact' equivalent circuits, which include the T and π equivalents derived from the coupled-circuit equations.

(a) If instead of a simple ratio changer a complex ratio changer is used, it is possible to produce an equivalent circuit having only three elements without using 'referred' values for the voltages and currents.[4,5]
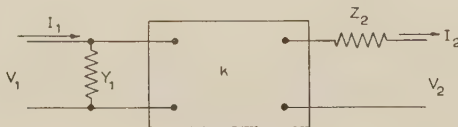
Fig. 6

For this the general circuit constants are given by (Fig. 6):

$$[O] = \begin{bmatrix} 1 & 0 \\ Y_1 & 1 \end{bmatrix} \begin{bmatrix} k & 0 \\ 0 & \frac{1}{k} \end{bmatrix} \begin{bmatrix} 1 & Z_2 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} k & kZ_2 \\ kY_1 & kZ_2 Y_1 + \frac{1}{k} \end{bmatrix}$$

Given the values of A, B, C, D it is clear that

$$k = A \text{ (the open-circuit voltage ratio)}$$

$$Z_2 = \frac{B}{k} = \frac{B}{A}$$

$$Y_1 = \frac{C}{k} = \frac{C}{A}$$

This circuit is known as 'an exact equivalent circuit determined by test'[4] or the 'exact practical equivalent circuit'[5] since k, Y_1 and Z_2 can be derived from open-circuit and short-circuit tests.

From eqns. (1) and (2),

$$V_1 = AV_2 + BI_2$$
$$I_1 = CV_2 + DI_2$$

when $I_2 = 0$

$$A = \frac{V_1}{V_2} = k$$

$$\frac{C}{A} = \frac{I_1}{V_1} = Y_1$$

so that k is the complex voltage ratio and Y_1 the admittance when the primary side is excited and the secondary open-circuited.

Also, from eqns. (4) and (5),

$$V_2 = DV_1 - BI_1$$
$$I_2 = - CV_1 + AI_1$$

when $\qquad V_1 = 0 \quad \frac{B}{A} = -\frac{V_2}{I_2} = Z_2$

and Z_2 is the impedance when the primary side is short-circuited and the secondary side excited, the sign being accounted for by the fact that the input is now on the secondary side.

(b) Variations on this circuit are possible. The expedient of transferring the ratio changer can still be resorted to, Z_2' being inserted between Y_1 and k and Z_2 being removed. Z_2' will now be $k^2 Z_2$, where k is complex and the values of V_2' and I_2' will be changed in phase as well as magnitude:

$$[O] = \begin{bmatrix} 1 & 0 \\ Y_1 & 1 \end{bmatrix} \begin{bmatrix} 1 & k^2 Z_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k & 0 \\ 0 & \frac{1}{k} \end{bmatrix}$$

$$[S_2]' = \begin{bmatrix} k & 0 \\ 0 & \frac{1}{k} \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}$$

$$[O]' = \begin{bmatrix} 1 & 0 \\ Y_1 & 1 \end{bmatrix} \begin{bmatrix} 1 & Z_2' \\ 0 & 1 \end{bmatrix}$$

(c) If the circuit shown in Fig. 6 is turned round, so that the impedance precedes the ratio changer and the admittance follows it, similar relations will hold, the only difference being that Z_1 is the impedance with the primary side excited and the secondary short-circuited, and Y_2 is the admittance with the primary side open-circuited and the secondary excited. k is the open-circuit voltage ratio as before.

(d) Again, the ratio changer may be transferred to the right so that both the impedance (Z_1) and the admittance (now Y_2/k^2) precede the changer. This, with the approximation of using k, the simple turns ratio, is the 'practical equivalent circuit' described by Rothe.[9]

All these equivalent circuits have the advantage over the common equivalent circuit that the values of the components can be determined from tests on the equipment.

They do not, however, present a physical picture of the transformer, and hence have little or no advantage over the 'abstract' 4-terminal network which can be used in their place.

A little consideration of the practical tests referred to in all these cases will show that they comprise:

(a) Measurement of an open-circuit admittance (or impedance).
(b) Measurement of a short-circuit impedance (or admittance).
(c) Measurement of the open-circuit voltage ratio (k, A) or more often $(k, |A|)$.

These then usually give two complex quantities and one scalar, so that there is actually insufficient information available to enable a full set of general circuit constants to be obtained and some approximation has to be used—such as that of neglecting the small phase angle of the general circuit constant A.

If it is assumed that the 'referred-to-primary' T-network is symmetrical, however, there is sufficient information—two complex quantities for components and one scalar for the 'reference'.

(4) THE INDUCTION MOTOR

(4.1) The Standard Equivalent Circuit

The exact equivalent circuit of the induction motor has the same form as that of the power transformer terminated in a load resistance representing the mechanical output from the rotor (Fig. 7).
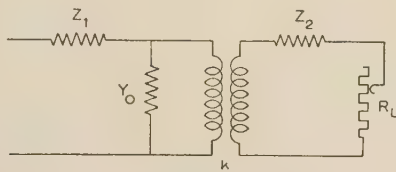
Fig. 7

It is usual to refer both Z_2 and R_L to the primary or stator so that the ratio changer is removed from further consideration.

In the approximate equivalent circuit the exciting admittance Y_0 is moved to the left of Z_1 just as in the transformer. For the induction motor the error caused by so doing will be greater than for the transformer since the exciting current is greater in proportion to the full-load current.

For both of these circuits the 4-terminal network method of approach may be used. As before, the usual equivalent circuit is only one out of many that are possible, but so far only one other equivalent circuit appears to have been proposed.[6]

(4.2) Determination of Circuit Elements from Test

The induction motor equivalent circuit can be considered as a 4-terminal network of the transformer type terminated by a resistance which represents the rotor output.

Since this resistance has no physical existence, the information which can be obtained from the usual running light and blocked-rotor tests is less complete than that obtainable from the open-circuit and short-circuit tests on the transformer to which they correspond.

In each test it is possible to determine the input impedance:

$$Z_{IN} = \frac{V_1}{I_1} = \frac{AR_L + B}{CR_L + D} \text{ [see eqn. (10)]}$$

(a) The open-circuit test is performed with the rotor running at synchronous speed when $R_L = \infty$ and $Z_{IN} = A/C$, or more usually with the rotor running light, in which case R_L is suffi-

ciently large for the test to correspond closely to a true open-circuit test.

(b) The short-circuit test is performed with the rotor blocked, when $R_L = 0$ and $Z_{IN} = B/D$.

From these two tests then only two quantities can be determined, and in order to specify a general 4-terminal network three are required—unless the network is assumed to be symmetrical or of the approximate equivalent type (L-network or cantilever).

For the squirrel-cage machine some such assumption is necessary; but for the wound-rotor machine more information is available from the standard tests, and another test is also possible, so that the three quantities required can be obtained.[6]

(4.3) The Wound-Rotor Machine

(a) In the short-circuit rotor-blocked test it is possible to determine, in addition to an input impedance, a transfer impedance $V_1/I_2 = B$, or a complex current ratio $I_1/I_2 = D$.

(b) If the rotor is open-circuited and held stationary it is possible to determine the open-circuit complex voltage ratio (A).

(c) If the machine is inverted, the open-circuit and short-circuit tests can be performed from the rotor side and also the test described in (b). From the inverted open-circuit test D/C is found, and from the inverted short-circuit test B/A [eqns. (3) and (4) being applied as in Section 3.3].

All the information necessary for determining the general circuit constants can therefore be obtained.

The exact practical circuit of Morris[6] uses the standard open-circuit test together with the rotor open-circuited test of (b) above and the inverted short-circuit test of (c) above.

(5) CIRCLE DIAGRAMS

The 4-terminal network and general circuit constant method of approach shows to advantage when circle diagrams are considered. It is not difficult to show that, when a passive linear 4-terminal network is terminated in an impedance which varies in such a manner that its impedance locus is a straight line, the input admittance locus is a circle.[8]

This application of the general bilinear transformation can readily be seen from the following analysis:

$$Z_{IN} = \frac{V_1}{I_1} = \frac{AZ_L + B}{CZ_L + D} \text{ (eqn. 10)}$$

and

$$Y_{IN} = \frac{CZ_L + D}{AZ_L + B}$$

Put $\quad Z = AZ_L + B \left(Z_L = \frac{Z - B}{A} \right)$ and $Y = \frac{1}{Z}$

$$Y_{IN} = \frac{\frac{C}{A}(Z - B) + D}{Z}$$

$$= \frac{CZ + (AD - BC)}{AZ} = \frac{C}{A} + \frac{1}{AZ}$$

$$= \frac{C}{A} + \frac{Y}{A} \text{ (since } AD - BC = 1)$$

Suppose Z_L to vary in such a way that its locus on an impedance diagram is a straight line. Then $Z = AZ_L + B$ implies a shift of the origin and a rotation and change in length of the Z_L locus, which is also straight line.

$Y = 1/Z$ will appear as a circular locus since in all practical cases the locus of Z does not pass through the origin.

$Y_{IN} = C/A + Y/A$ will also appear as a circular locus (the

origin being shifted and the location and diameter of the circle being altered).[7,8]

It should be apparent that the input admittance circle diagram is quite independent of the form of the particular equivalent circuit chosen and is dependent only on the nature of the terminating impedance.

An obvious application arises for the induction motor in which the common transformer-type equivalent circuit is used and the load is represented by a variable resistance—thus satisfying the straight-line impedance locus requirement.

It will apply to any machine provided that it can be represented with sufficient accuracy by a passive linear 4-terminal network.

It has been shown by the author[7] that the circle diagram which corresponds to the case of the four-terminal network

$$[O] = \begin{bmatrix} A \underline{/\alpha} & B \underline{/\beta} \\ C \underline{/\gamma} & D \underline{/\delta} \end{bmatrix}$$

terminated by an impedance $Z_L = Z_L \underline{/\zeta}$, where Z_L is variable and ζ constant, is as shown in Fig. 8,
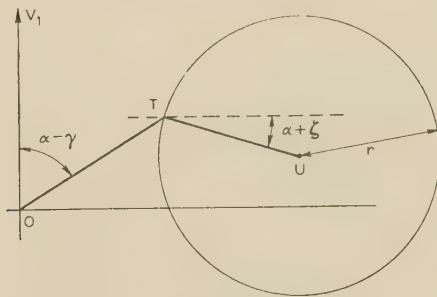
Fig. 8

where $OT = V_1\left(\dfrac{C}{A}\right)$, $UT = r = \dfrac{V_1}{2AB}\sqrt{\left(\dfrac{n^2 + 1}{n^2}\right)}$

and $n = \tan(\alpha + \zeta - \beta)$

$V_1 = |V_1|$

[For the induction motor $(2\alpha + \zeta)$ is small and negative.]

(6) CONCLUSION

If a physical picture of the apparatus represented is not required the 4-terminal network presents advantages as an equivalent circuit. It is useful in designing the tests to be carried out on the equipment and in indicating the information to be expected from them.

The use of simple matrix algebra enables the general circuit constants of given equivalent circuits to be determined and also enables an assessment of the degree of approximation involved to be made.

(7) REFERENCES

(1) VOWELS, R.: 'Matrix Methods in the Solution of Ladder Networks', *Journal I.E.E.*, 1948, **95**, Part III, p. 40.

(2) GUERTLER, R.: 'An Introduction to the Matrix Method of Solving Electrical Network Problems', *Journal of the Institution of Engineers, Australia*, 1950, **22**, p. 46.

(3) Central Station Engineers, Westinghouse Electric and Manufacturing Company: 'Electrical Transmission and Distribution Reference Book' (Westinghouse Electric and Manufacturing Co., July, 1944), pp. 97, 98.

(4) Electrical Engineering Staff, Massachusetts Institute of Technology: 'Magnetic Circuits and Transformers' (John Wiley, New York, 1943), pp. 352, 355, 357 and 455.

(5) MORRIS, D.: 'Some Tests of an Exact Practical Theory of the Transformer', *Proceedings I.E.E.*, Paper No. 902 S, February, 1950 (**97**, Part II, p. 17. Also Discussion, p. 735).

(6) MORRIS, D.: 'Some Tests of an Exact Practical Theory of the Induction Motor', *ibid.*, Paper No. 1041 U, December, 1950 (**97**, Part II, p. 767).

(7) MOORHOUSE, C. E.: 'Classes of Circle Diagrams', *Journal of the Institution of Engineers, Australia*, 1950, **22**, p. 69.

(8) TARBOUX, J. G.: 'A Generalized Circle Diagram for a Four Terminal Network and its Application to the Capacitor Single Phase Motor', *Transactions of the American I.E.E.*, 1954, **64**, p. 881.

(9) ROTHE, F. S.: 'An Introduction to Circuit Analysis' (John Wiley, New York, 1953), pp. 1–6.

A NEW THEOREM IN ELECTROSTATICS WITH APPLICATIONS TO CALCULABLE STANDARDS OF CAPACITANCE

By D. G. LAMPARD, M.Sc., Ph.D.

SUMMARY

The paper presents what is believed to be a new theorem in electrostatics, showing that the direct capacitance of a very general type of cylindrical 3-terminal capacitor is a constant, namely $(\log_\varepsilon 2)/4\pi^2$.

The presentation and proof of this theorem is followed by a detailed analysis of a cylindrical 3-terminal capacitor of rectangular cross-section, in which the effects of various practical limitations are evaluated in quantitative terms. The results of similar analyses of cylindrical 3-terminal capacitors of other cross-sections are stated.

As a consequence of this work, it appears that such cylindrical 3-terminal capacitors are extremely attractive as precise calculable standards of capacitance, only one precision length measurement being necessary to enable their capacitance to be computed to a high order of accuracy.

(1) INTRODUCTION

The practical system of electrical units is based on electromagnetic laws, and the corresponding standards of measurement are established by reference to some form of inductor whose value in electromagnetic units can be determined from its mechanical dimensions. A calculable capacitor, on the other hand, yields an electrostatic unit which is related to the electromagnetic unit by c, the velocity of light. The precision with which c is known has increased considerably in recent years and makes possible the use of a sufficiently precise capacitor as an alternative means of establishing the electrical standards of measurement.

Several types of capacitor, notably those based on concentric cylinders,[1] parallel plates[2] and modifications[3,4] of these, have been described, but despite the fact that capacitance has the dimensions of length only, these types all suffer from the disadvantage that several precision length measurements must be made to enable the capacitance to be computed to a high order of accuracy.*

Attention has therefore been directed to cylindrical systems, since only ratios of dimensions are required to determine the shape of the cross-section. Since unity is a ratio which is easily established experimentally, the properties of a cylindrical capacitor of *square* cross-section were examined theoretically. It was found that if a square cylinder is constructed from four conducting planes which are insulated from one another at the corners, then the direct capacitances per unit length of cylinder, between each pair of opposing inside faces, are equal and independent of the size of the square. This direct capacitance was calculated as $(\log_\varepsilon 2)/4\pi^2$ cm/cm, and it was also calculated that if the cross-section is rectangular, with a ratio of sides of $(1 + \delta):1$, where $\delta \ll 1$, then the mean of the two capacitances is

$$\frac{\log_\varepsilon 2}{4\pi^2}(1 + 3 \cdot 45\delta^2 + \text{higher terms}) \text{ cm/cm}$$

Correspondence on Monographs is invited for consideration with a view to publication.

Dr. Lampard is in the Division of Electrotechnology of the Commonwealth Scientific and Industrial Research Organization, New South Wales, Australia.

A rough estimate of the effect of the necessarily finite insulating gaps at the corners of the square was obtained by computing the integrated charge distribution function in the vicinity of the corners. It was shown that the fraction $D(\xi)$ of the total charge, which lies in a pair of small end strips each of fractional width ξ, where $\xi \ll 1$, is given by

$$D(\xi) = 4 \cdot 9596\xi^2 + \text{higher terms}$$

These calculations confirm that, provided that asymmetry and gap effects are not too great, a standard capacitor could be constructed which requires only one length measurement for the computation of its capacitance. The length of the capacitor may be defined by insulating a length of one face from the remainder of that face, the remainder functioning as a guard.

The capacitance per unit length and asymmetry effects were calculated for a number of other symmetrical cross-sections. These calculations all gave the same value, namely $(\log_\varepsilon 2)/4\pi^2$ e.s.u., for the capacitance per unit length, and it was shown that this result was a direct consequence of a rather general theorem.

A preliminary note which briefly summarizes some of this work has already been published.[6] The present paper describes these investigations in some detail, and in Section 2 the general theorem mentioned above is stated and proved. Section 3 is devoted to the discussion of a cylindrical capacitor of rectangular cross-section, and, in particular, the effects of both finite gaps and departures from symmetry are examined. Results for cylindrical capacitors of other cross-sections, corresponding to those given in Section 3, are listed in Section 4. General results, concerning the effects of deviations from symmetry, of finite gap widths and of dielectric films on the electrode surfaces, are obtained as corollaries of some of these examples.

(2) A GENERAL THEOREM ON CYLINDRICAL CAPACITORS

Let the closed curve S (see Fig. 1) be the right cross-section of a conducting cylindrical shell, the cross-section having one axis of symmetry AC, but being otherwise arbitrary. Further, let this shell be divided into four parts by two planes at right angles, the line of intersection of the planes being parallel to the gene-
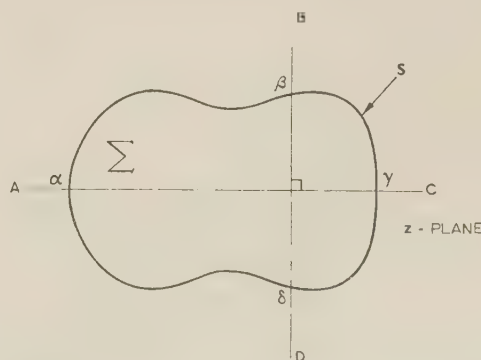
Fig. 1.—Cross-section of cylindrical capacitor.

rators of the cylinder, and one of the planes containing the line AC.

Then, the direct capacitance per unit length of the cylinder, between opposing parts of the shell (e.g. $\alpha\delta$ to $\beta\gamma$), due to the field inside (or outside) the shell, is a constant:

$$C_0 = \frac{\log_\varepsilon 2}{4\pi^2} \text{ e.s.u.} = 0\cdot0175576 \text{ e.s.u.} \quad . \quad . \quad (1)$$

Proof.

An existence theorem (see Reference 7, p. 186) due to Riemann states:

> If Σ is the open domain bounded by a simple closed Jordan curve S, there *exists* a unique analytic function f(z) regular in Σ, such that $\omega = $ f(z) maps Σ conformally on $|\omega| < 1$, and also transforms a given point $z = z_0$ within S into the origin, and a given direction at $z = z_0$ into the positive direction of the real axis.

Referring to Fig. 1, any point is now chosen which is inside S and which lies on AC, as the point $z = z_0$. Applying Riemann's theorem, it is seen that the cross-section of the cylindrical shell of Fig. 1 may be transformed into a circle, as shown in Fig. 2,
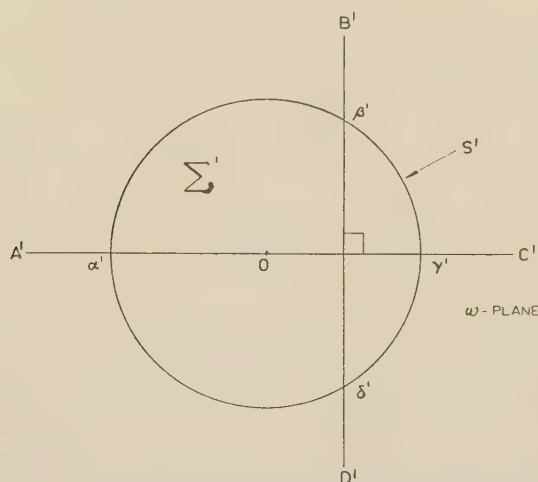
Fig. 2.—Cross-section of cylindrical capacitor after conformal transformation.

in such a way that the line AC maps into the line A′C′, and that symmetry is preserved.

The direct capacitance per unit length of cylinder between the opposing parts, say $\alpha'\delta'$ and $\beta'\gamma'$, due to the field *inside* the cylindrical shell of Fig. 2 can now be computed. The computation for the capacitance due to the field *outside* the cylindrical shell is similar and details are therefore omitted.

It is convenient first to redraw Fig. 2, at the same time adopting a system of polar co-ordinates with the line OX (see Fig. 3) as origin for the angular co-ordinate.

By definition, the direct capacitance per unit length of cylinder between the faces $\beta'\gamma'$ and $\alpha'\delta'$ is simply the total charge per unit length of the cylinder which appears on the face $\alpha'\delta'$, all faces of the cylinder being at zero potential with the exception of $\beta'\gamma'$, which is at unit potential. This direct capacitance is simply Maxwell's 'coefficient of induction' with sign reversed (see Reference 9 and Reference 8, p. 38).

Thus, if $V(r, \phi)$ denotes the potential function inside the cylinder (whose cross-section is circular with radius a, say), it is clear that the following boundary conditions must be imposed:

$$V(a, \phi) = 1, \quad -\frac{\theta}{2} \leqslant \phi \leqslant +\frac{\theta}{2} \quad . \quad . \quad . \quad (2a)$$

$$= 0, \quad \frac{\theta}{2} < \phi < 2\pi - \frac{\theta}{2} \quad . \quad . \quad . \quad (2b)$$
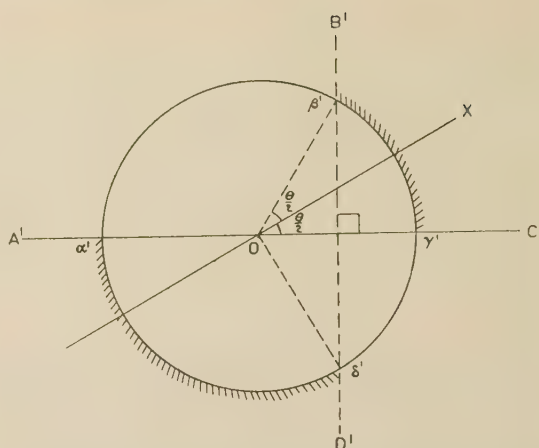
Fig. 3.—Transformed cross-section of cylindrical capacitor showing co-ordinate system used in analysis.

At all points inside the shell, $V(r, \phi)$ must satisfy Laplace's equation,[8] namely

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r}\frac{\partial V}{\partial r} + \frac{1}{r^2}\frac{\partial^2 V}{\partial \phi^2} = 0 \quad . \quad . \quad . \quad (3)$$

It is further required that $V(r, \phi)$

(a) be an *even*, periodic function of ϕ, because of the choice of OX as origin for the angular co-ordinate,
(b) have no singularity at $r = 0$.

It is easy to show that the most general solution of eqn. (3) which satisfies these requirements is of the form

$$V(r, \phi) = \sum_{m=0}^{\infty} A(m)r^m \cos m\phi \quad . \quad . \quad . \quad (4)$$

When we set $r = a$, we have

$$V(a, \phi) = \sum_{m=0}^{\infty} A(m)a^m \cos m\phi \quad . \quad . \quad . \quad (5)$$

which is a Fourier cosine series in ϕ, and, using the boundary conditions (2), we find

$$A(m)a^m = \frac{1}{\pi}\int_{-\theta/2}^{+\theta/2} \cos m\phi d\phi; \quad m = 1, 2, 3 \quad . \quad . \quad (6)$$

$$= \frac{2}{m\pi}\sin\frac{m\theta}{2}; \quad m = 1, 2, 3 \quad . \quad . \quad (7)$$

$$A(0) = \frac{\theta}{2\pi} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

so that, finally,

$$V(r, \phi) = \frac{\theta}{2\pi} + \sum_{m-1}^{\infty} \frac{2}{m\pi}\left(\frac{r}{a}\right)^m \sin\frac{m\theta}{2} \cos m\phi \quad . \quad (9)$$

The charge density per unit length at the surface, $\sigma(\phi)$, is obtained by application of Gauss's theorem (see Reference 9, p. 17) and is given by

$$\sigma(\phi) = -\frac{1}{4\pi}\frac{\partial V}{\partial r}\bigg]_{r=a} \quad . \quad . \quad . \quad . \quad (10)$$

$$= -\frac{1}{2\pi^2 a}\sum_{m=1}^{\infty} \sin\frac{m\theta}{2}\cos m\phi \quad . \quad . \quad (11)$$

The series on the right-hand side of eqn. (11) is easily trans-

formed into a pair of geometric series of complex exponentials, which can be summed to give

$$\sigma(\phi) = -\frac{1}{8\pi^2 a}\left[\frac{\sin(\theta/2 + \phi)}{1 - \cos(\theta/2 + \phi)} + \frac{\sin(\theta/2 - \phi)}{1 - \cos(\theta/2 - \phi)}\right] \quad (12)$$

The capacitance per unit length of the cylinder is now obtained by integrating $\sigma(\phi)$ over the face $\alpha'\delta'$.

Thus, using eqn. (12), we find

$$\int_{-\theta/2}^{2\pi - 3\theta/2} \sigma(\phi)a d\phi = -\frac{1}{8\pi^2}[\log_\varepsilon(1 - \cos\theta) - \log_\varepsilon(1 - \cos 2\theta)$$
$$- \log_\varepsilon 2 + \log_\varepsilon(1 + \cos\theta)] \quad (13)$$
$$= \frac{1}{4\pi^2}\log_\varepsilon 2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Remembering that capacitance is invariant under the transformation used here, it can be seen that eqn. (14) establishes the proof of the theorem.

(3) ANALYSIS OF A CYLINDRICAL CAPACITOR OF RECTANGULAR CROSS-SECTION

In this Section a detailed analysis is given of a cylindrical capacitor whose cross-section is a rectangle, as shown in Fig. 4.
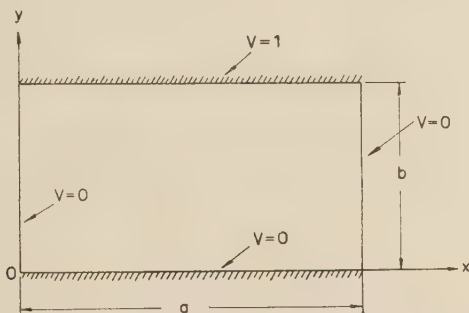
Fig. 4.—Cross-section and co-ordinate system for cylindrical capacitor of rectangular cross-section.

Of special interest will be the case in which the rectangle becomes a square, when the direct capacitance per unit length will be given immediately by the theorem of the preceding Section. However, since deviations from squareness would be an important practical perturbation, the more general rectangular case is treated.

By definition, the direct capacitance per unit length between the faces $y = 0$ and $y = b$ is the total charge per unit length which appears on the face $y = 0$, when all faces are maintained at zero potential except the face $y = b$, which is raised to unit potential.

If $V(x, y)$ denotes the potential function inside the rectangle, it is clear, from the above discussion, that the following boundary conditions must be imposed:

$$V(0, y) = 0, \quad 0 \leqslant y \leqslant b \quad . \quad . \quad . \quad (15a)$$
$$V(x, 0) = 0, \quad 0 \leqslant x \leqslant a \quad . \quad . \quad . \quad (15b)$$
$$V(a, y) = 0, \quad 0 \leqslant y \leqslant b \quad . \quad . \quad . \quad (15c)$$
$$V(x, b) = 1, \quad 0 \leqslant x \leqslant a \quad . \quad . \quad . \quad (15d)$$

The most general solution of Laplace's equation which satisfies boundary conditions (15a), (15b) and (15c) is of the form

$$V(x, y) = \sum_{n=1}^{\infty} A(n)\sin\frac{n\pi x}{a}\sinh\frac{n\pi y}{a} \quad . \quad . \quad . \quad (16)$$

Thus, from the theory of Fourier series, it follows that, after setting $y = b$, and making use of condition (15d),

$$A(n)\sinh\frac{n\pi b}{a} = \frac{2}{a}\int_0^a \sin\frac{n\pi x}{a}dx \quad . \quad . \quad . \quad (17)$$

$$= \frac{2}{n\pi}[1 - (-1)^n] \quad . \quad . \quad . \quad (18)$$

so that the complete solution is

$$V(x, y) = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{[1 - (-1)^n]}{n}\frac{\sinh\frac{\pi n y}{a}}{\sinh\frac{\pi n b}{a}}\sin\frac{\pi n x}{a} \quad (19)$$

It follows from Gauss's theorem that $\sigma(x)$, the surface charge density per unit length on the face $y = 0$, is given by

$$\sigma(x) = \frac{1}{4\pi}\frac{\partial}{\partial y}V(x, y)\bigg]_{y=0} \quad . \quad . \quad . \quad . \quad (20)$$

$$= \frac{1}{2\pi a}\sum_{n=1}^{\infty}[1 - (-1)^n]\frac{\sin\frac{\pi n x}{a}}{\sinh\frac{\pi n b}{a}} \quad . \quad . \quad (21)$$

Integration of $\sigma(x)$ over the face $y = 0$ gives, finally, the capacitance per unit length between the faces $y = 0$ and $y = b$ as

$$C(\beta) = \int_0^a \sigma(x)dx \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

$$= \frac{2}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n + 1)\sinh(2n + 1)\pi\beta} \quad . \quad (23)$$

where

$$\beta = b/a \quad . \quad . \quad . \quad . \quad . \quad (24)$$

In Section 8 it is shown that the series on the right-hand side of eqn. (23) may be summed to give

$$C(\beta) = \frac{1}{\pi^2}\log_\varepsilon\frac{\theta_3(0, \varepsilon^{-\pi\beta})}{\theta_4(0, \varepsilon^{-\pi\beta})} \quad . \quad . \quad . \quad (25)$$

where

$$\theta_3(0; q) = 1 + 2\sum_{n=1}^{\infty}q^{n^2} = \left[\frac{2}{\pi}K(k)\right]^{1/2} \quad . \quad . \quad (26)$$

$$\theta_4(0; q) = 1 + 2\sum_{n=1}^{\infty}(-1)^n q^{n^2} = \left[\frac{2}{\pi}(1 - k^2)^{1/2}K(k)\right]^{1/2} \quad (27)$$

are theta functions (see Reference 10, p. 464), and $K(k)$ is the complete elliptic integral of the first kind* with modulus k chosen so that

$$\frac{K[\sqrt{(1 - k^2)}]}{K(k)} = \frac{1}{\pi}\log_\varepsilon\left(\frac{1}{q}\right) \quad . \quad . \quad . \quad (28)$$

From eqns. (26), (27) and (28), we have

$$C(\beta) = \frac{1}{4\pi^2}\log_\varepsilon\left(\frac{1}{1 - k^2}\right) \quad . \quad . \quad . \quad (29)$$

where k satisfies

$$\frac{K[\sqrt{(1 - k^2)}]}{K(k)} = \beta \quad . \quad . \quad . \quad . \quad . \quad (30)$$

When it is recalled that $C(\beta)$ denotes the direct capacitance per unit length between the faces $y = 0$ and $y = b$ with $\beta = b/a$,

* The theta functions and elliptic integrals which occur here arise quite naturally as a result of applying the Schwarz–Christoffel transformation to a rectangle. Some very general results concerning the solution of Laplace's equation inside a rectangular boundary have been given in a recent paper.[11]

it can be seen that $C\left(\frac{1}{\beta}\right)$ is the direct capacitance per unit length between the faces $x = 0$ and $x = a$.

It follows from eqns. (29) and (30) that

$$C\left(\frac{1}{\beta}\right) = \frac{1}{4\pi^2} \log_\varepsilon \frac{1}{k^2} \quad \ldots \ldots \quad (31)$$

and hence, again making use of eqn. (30), $C(\beta)$ and $C\left(\frac{1}{\beta}\right)$ are connected by the identity

$$\varepsilon^{-4\pi^2 C(\beta)} + \varepsilon^{-4\pi^2 C\left(\frac{1}{\beta}\right)} = 1 \quad \ldots \ldots \quad (32)$$

In particular, when $\beta = 1$ the rectangle becomes a square for which the diagonal is an axis of symmetry, and it follows immediately from eqn. (32) that

$$C(1) = \frac{1}{4\pi^2} \log_\varepsilon 2 \quad \ldots \ldots \quad (33)$$

which is in agreement with the theorem given in Section 1.

Explicit solutions of eqn. (30) are known (see Reference 10, p. 525) for some other values of β. [The solution for $\beta = 1$ is closely related to the theory of lemniscate functions (see Reference 10, p. 524).] These solutions, together with corresponding values of $C(\beta)$ and $C\left(\frac{1}{\beta}\right)$ obtained by substituting these solutions in eqns. (29) and (31), are given in Table 1.

The behaviour of the functions $C(\beta)$ and $C\left(\frac{1}{\beta}\right)$ and their arithmetic mean is shown in Fig. 5 for $\frac{1}{2} \leqslant \beta \leqslant 2$. It will be seen that the mean capacitance per unit length has a mini-
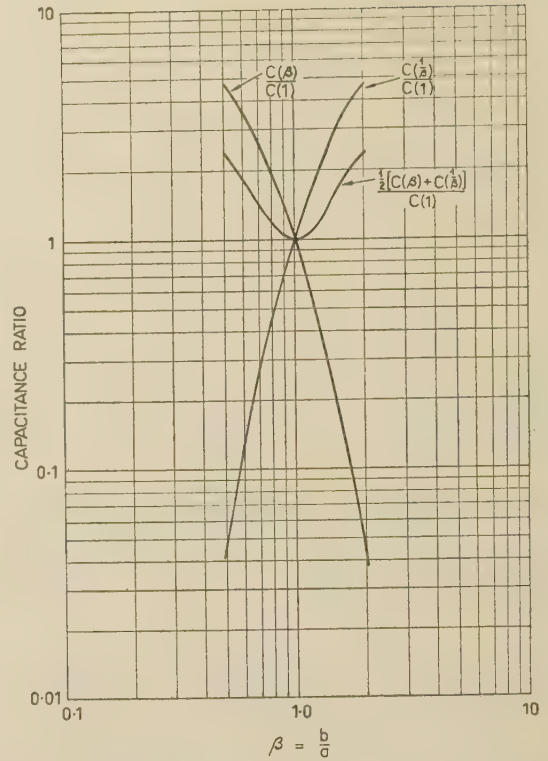
Fig. 5.—Cylindrical capacitor: rectangular cross-section.

Table 1

β	k	$C(\beta)$	$C\left(\frac{1}{\beta}\right)$	$\frac{1}{2}\left[C(\beta) + C\left(\frac{1}{\beta}\right)\right]$
1	$\frac{1}{\sqrt{2}}$	$\frac{1}{4\pi^2} \log_\varepsilon 2$	$\frac{1}{4\pi^2} \log_\varepsilon 2$	$\frac{1}{4\pi^2} \log_\varepsilon 2$
$\sqrt{2}$	$\sqrt{(2)} - 1$	$\frac{1}{4\pi^2} \log_\varepsilon \frac{\sqrt{(2)}+1}{2}$	$\frac{1}{4\pi^2} \log_\varepsilon \frac{1}{3-2\sqrt{2}}$	$\frac{1}{4\pi^2} \frac{1}{2} \log_\varepsilon \frac{7+5\sqrt{2}}{2}$
$\sqrt{3}$	$\frac{1}{2}(2-\sqrt{3})^{1/2}$	$\frac{1}{4\pi^2} \log_\varepsilon \frac{4}{2+\sqrt{3}}$	$\frac{1}{4\pi^2} \log_\varepsilon \frac{4}{2-\sqrt{3}}$	$\frac{1}{4\pi^2} 2 \log_\varepsilon 2$
2	$\frac{\sqrt{2}-1}{\sqrt{2}+1}$	$\frac{1}{4\pi^2} \log_\varepsilon \frac{4+3\sqrt{2}}{8}$	$\frac{1}{4\pi^2} \log_\varepsilon (17+12\sqrt{2})$	$\frac{1}{4\pi^2} \frac{1}{2} \log_\varepsilon \frac{140+99\sqrt{2}}{2}$

mum at $\beta = 1$, showing that, to the first order, small departures from squareness produce no change in the mean capacitance per unit length. Before taking up this point in more detail, the asymptotic $(\beta \to \infty)$ behaviour of $C(\beta)$ and $C\left(\frac{1}{\beta}\right)$ is investigated. Returning to expression (25) for $C(\beta)$, and making use of the series definitions (26) and (27) for the theta functions, we obtain

$$C(\beta) = \frac{1}{\pi^2}\left(4\varepsilon^{-\pi\beta} + \frac{16}{3}\varepsilon^{-3\pi\beta} - 8\varepsilon^{-5\pi\beta} + 20\varepsilon^{-9\pi\beta} + \ldots\right) \quad (34)$$

$$\sim \frac{4}{\pi^2}\varepsilon^{-\pi\beta} \text{ as } \beta \to \infty \quad \ldots \ldots \ldots \quad (35)$$

where the Taylor expansion has been used for the logarithm

occurring in eqn. (25). The corresponding expression for $C\left(\frac{1}{\beta}\right)$ is obtained by making use of the identity (32), thus:

$$C\left(\frac{1}{\beta}\right) = \frac{1}{4\pi^2} \log_\varepsilon \left[1 - \varepsilon^{-4\pi^2 C(\beta)}\right]^{-1} \quad \ldots \quad (36)$$

$$\sim \frac{1}{4\pi^2} \log_\varepsilon \left(\frac{\varepsilon^{\pi\beta}}{16}\right) \text{ as } \beta \to \infty \quad \ldots \quad (37)$$

$$\sim \frac{\beta}{4\pi} \text{ as } \beta \to \infty \quad \ldots \ldots \quad (38)$$

We note that the results in eqns. (35) and (38) may readily be obtained by working directly with expression (23) for $C(\beta)$.

To see the way in which small deviations from squareness

affect the mean capacitance per unit length, we use eqn. (34) and put

$$\beta = 1 + \delta, \text{ where } \delta \ll 1 \quad . \quad . \quad . \quad . \quad (39)$$

We then have

$$C(1 + \delta) = \frac{1}{\pi^2}\left[4\varepsilon^{-\pi(1+\delta)} + \frac{16}{3}\varepsilon^{-3\pi(1+\delta)} \right.$$
$$\left. - 8\varepsilon^{-5\pi(1+\delta)} + 20\varepsilon^{-9\pi(1+\delta)} + \ldots \right] \quad (40)$$

$$= C(1) - \frac{\delta}{\pi}(4\varepsilon^{-\pi} + 16\varepsilon^{-3\pi} - 40\varepsilon^{-5\pi} + 180\varepsilon^{-9\pi} + \ldots)$$

$$+ \frac{\delta^2}{2}(4\varepsilon^{-\pi} + 48\varepsilon^{-3\pi} - 200\varepsilon^{-5\pi} + 1\,620\varepsilon^{-9\pi} + \ldots)$$

$$+ \text{ terms in } \delta^3 \text{ and higher} \quad . \quad . \quad . \quad (41)$$

Similarly we find

$$C\left(\frac{1}{1 + \delta}\right) =$$

$$C(1) + \frac{\delta}{\pi}(4\varepsilon^{-\pi} + 16\varepsilon^{-3\pi} - 40\varepsilon^{-5\pi} + 180\varepsilon^{-9\pi} + \ldots$$

$$+ \frac{\delta^2}{2}\left[\left(1 - \frac{2}{\pi}\right)4\varepsilon^{-\pi} + 16\left(3 - \frac{2}{\pi}\right)\varepsilon^{-3\pi}\right.$$

$$\left. - 40\left(5 - \frac{2}{\pi}\right)\varepsilon^{-5\pi} + \ldots\right] + \text{ terms in } \delta^3 \text{ and higher} \quad (42)$$

Adding eqns. (41) and (42) we obtain finally

$$\frac{1}{2}\left[C(1 + \delta) + C\left(\frac{1}{1 + \delta}\right)\right] =$$

$$C(1) + \frac{\delta^2}{2}\left[\left(1 - \frac{1}{\pi}\right)4e^{-\pi} + \left(3 - \frac{1}{\pi}\right)16\varepsilon^{-3\pi}\right.$$

$$\left. - \left(5 - \frac{1}{\pi}\right)40\varepsilon^{-5\pi} + \ldots\right] + \text{ terms in } \delta^3 \text{ and higher} \quad (43)$$

$$= C(1)[1 + 3\cdot453\,5\delta^2 + \text{ terms in } \delta^3 \text{ and higher}] \quad . \quad (44)$$

which shows clearly that small deviations from squareness produce only a second-order change in the mean capacitance per unit length.

To complete our investigation of the cylindrical capacitor of rectangular cross-section, it would be desirable to have an estimate of the error introduced by the presence of the necessarily finite insulating gaps. An exact determination of this error is complicated, but an upper bound is readily obtained by computing the fraction of the total charge (on the face $y = 0$) which lies in a pair of narrow strips, each of fractional width ξ, lying at each end $x = 0$ and $x = a$ of the face $y = 0$. It is thought that this upper bound, while probably being considerably larger than the true error, is nevertheless a useful figure, by means of which gap effects in different types of capacitor may be roughly compared.

The surface charge density per unit length on the face $y = 0$ is given by eqn. (21), and the charge which lies in the strip $0 \leqslant x \leqslant \xi a$ is

$$\int_0^{\xi a} \sigma(x)dx = \frac{1}{2\pi a}\sum_{n=1}^{\infty}\frac{[1 - (-1)^n]}{\sinh \pi n\beta}\int_0^{\xi a}\sin\frac{\pi n x}{a}dx \quad . \quad (45)$$

where, as before, $\beta = \dfrac{b}{a}$.

i.e. $$\int_0^{\xi a} \sigma(x)dx = \frac{1}{\pi^2}\sum_{n=0}^{\infty}\frac{[1 - \cos(2n + 1)\pi\epsilon]}{(2n + 1)\sinh(2n + 1)\pi\beta} \quad . \quad (46)$$

By symmetry, a similar amount of charge lies in the strip $a(1 - \xi) \leqslant x \leqslant a$, so that, after expanding the cosine and noting that each of the resulting infinite series converges, the charge which lies in both strips is given by

$$2\int_0^{\xi a} \sigma(x)dx = \xi^2\sum_{n=0}^{\infty}\frac{2n + 1}{\sinh(2n + 1)\pi\beta} + \text{ terms in } \xi^4 \text{ and higher} \quad . \quad . \quad . \quad (47)$$

Now

$$\sum_{n=0}^{\infty}\frac{(2n + 1)}{\sinh(2n + 1)\pi\beta} = 2\sum_{s=0}^{\infty}\left[\sum_{n=0}^{\infty}(2n + 1)\varepsilon^{-(2s+1)(2n+1)\pi\beta}\right] \quad (48)$$

$$= -2\sum_{s=0}^{\infty}\frac{1}{\pi(2s + 1)}\left[\frac{d}{d\beta}\sum_{n=0}^{\infty}\varepsilon^{-(2s+1)(2n+1)\pi\beta}\right] \quad . \quad . \quad . \quad (49)$$

$$= -\frac{1}{\pi}\frac{d}{d\beta}\sum_{s=0}^{\infty}\frac{1}{(2s + 1)\sinh(2s + 1)\pi\beta} \quad (50)$$

$$= -\frac{\pi}{2}\frac{d}{d\beta}[C(\beta)] \quad . \quad . \quad . \quad . \quad (51)$$

Here we have made use of eqn. (23) and also of the identity

$$\frac{1}{\sinh x} = 2\sum_{s=0}^{\infty}\varepsilon^{-(2s+1)x} \quad . \quad . \quad . \quad . \quad (52)$$

Substituting eqn. (51) in eqn. (47), we find that

$$2\int_0^{\xi a} \sigma(x)dx = -\frac{\pi}{2}\xi^2 C'(\beta) + \text{ terms in } \xi^4 \text{ and higher} \quad (53)$$

where the prime denotes differentiation with respect to β.

The fraction $D(\xi)$ of the total charge which lies in the small gaps is then given by

$$D(\xi) = \frac{2\int_0^{\xi a}\rho(x)dx}{\int_0^a\rho(x)dx} = \xi^2\frac{\pi}{2}\left[-\frac{C'(\beta)}{C(\beta)}\right] + \text{ terms in } \xi^4 \text{ and higher} \quad . \quad . \quad . \quad (54)$$

where $C'(\beta)$ may be obtained by differentiating eqn. (23). A more rapidly converging series for numerical computation is obtained, however, by differentiating eqn. (25) after making use of the series definitions (26) and (27) for the theta functions. Using this method it is easily shown that

$$-C'(\beta) = \frac{1}{\pi}\left[\frac{2\sum_{n=1}^{\infty}n^2\varepsilon^{-\pi\beta n^2}}{1 + 2\sum_{n=1}^{\infty}\varepsilon^{-\pi\beta n^2}} - \frac{2\sum_{n=1}^{\infty}(-1)^n n^2\varepsilon^{-\pi\beta n^2}}{1 + 2\sum_{n=1}^{\infty}(-1)^n n^2\varepsilon^{-\pi\beta n^2}}\right] \quad . \quad . \quad . \quad (55)$$

In particular, when $\beta = 1$, substitution from eqn. (33) gives

$$D(\xi)_{\beta=1} = \xi^2\frac{4\pi^2}{\log_\varepsilon 2}\left(\frac{\varepsilon^{-\pi} + 4\varepsilon^{-4\pi} + 9\varepsilon^{-9\pi} + \ldots}{1 + 2\varepsilon^{-\pi} + 2\varepsilon^{-4\pi} + 2\varepsilon^{-9\pi} + \ldots}\right.$$

$$\left. + \frac{\varepsilon^{-\pi} - 4\varepsilon^{-4\pi} + 9\varepsilon^{-9\pi} - \ldots}{1 - 2\varepsilon^{-\pi} + 2\varepsilon^{-4\pi} - 2\varepsilon^{-9\pi} + \ldots}\right)$$

$$+ \text{ terms in } \xi^4 \text{ and higher} \quad (56)$$

$$= 4\cdot959\,8\xi^2 + \text{ terms in } \xi^4 \text{ and higher} \quad . \quad . \quad (57)$$

and we see that the fractional change in capacitance produced by the small but finite insulating gaps is a second-order effect. This property and those properties expressed by eqns. (33) and (44), show that the cylindrical capacitor of square cross-section would be an attractive design for a calculable capacitance standard.

(4) RESULTS FOR CYLINDRICAL CAPACITORS OF OTHER CROSS-SECTIONS

In this Section we list, for some cylindrical capacitors of other cross-sections, results corresponding to those given in Section 3 for the capacitor of rectangular cross-section.

(4.1) Circular Cylinder

Using the notation of Fig. 6 and taking OX as the origin for the angular co-ordinate we have:

Potential at the point $(r; \phi)$.

$$V(r; \phi) = \frac{\theta_1}{2\pi} + \sum_{m=1}^{\infty} \frac{2}{m\pi} \left(\frac{r}{a}\right)^m \sin \frac{m\theta_1}{2} \cos m\phi, \quad r \leqslant a \quad (58)$$
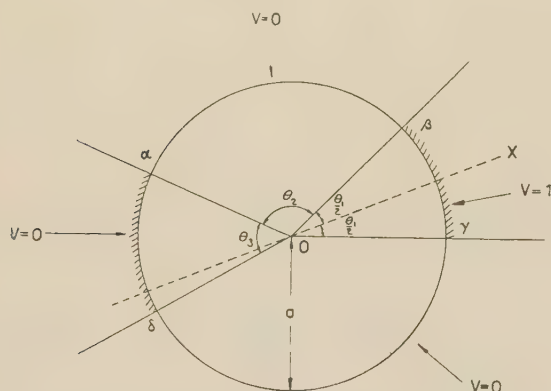
Fig. 6.—Cross-section and co-ordinate system for cylindrical capacitor of circular cross-section (general case).

Surface charge density per unit length.

$$\sigma(\phi) = -\frac{1}{8\pi^2 a} \left[\frac{\sin\left(\frac{\theta_1}{2} + \phi\right)}{1 - \cos\left(\frac{\theta_1}{2} + \phi\right)} + \frac{\sin\left(\frac{\theta_1}{2} - \phi\right)}{1 - \cos\left(\frac{\theta_1}{2} - \phi\right)} \right] \quad (59)$$

Capacitance per unit length between hatched portions.

$$C_1(\theta_1, \theta_2, \theta_3) = \frac{1}{8\pi^2} \log_\varepsilon \frac{[1 - \cos(\theta_1 + \theta_2)][1 - \cos(\theta_2 + \theta_3)]}{[1 - \cos(\theta_1 + \theta_2 + \theta_3)][1 - \cos\theta_2]} \quad (60)$$

Capacitance per unit length between unhatched portions.

$$C_2(\theta_1, \theta_2, \theta_3) = \frac{1}{8\pi^2} \log_\varepsilon \frac{[1 - \cos(\theta_1 + \theta_2)][1 - \cos(\theta_2 + \theta_3)]}{[1 - \cos\theta_1][1 - \cos\theta_3]} \quad (61)$$

Mean capacitance per unit length.

$$C_M(\theta_1, \theta_2, \theta_3) = \frac{1}{2}(C_1 + C_2) = \frac{1}{8\pi^2} \log_\varepsilon$$

$$\left\{ \frac{[1 - \cos(\theta_1 + \theta_2)]^2[1 - \cos(\theta_2 + \theta_3)]^2}{(1 - \cos\theta_1)(1 - \cos\theta_2)(1 - \cos\theta_3)[1 - \cos(\theta_1 + \theta_2 + \theta_3)]} \right\} \quad (62)$$

Fraction $D(\xi)$ of total charge which lies in the gaps β, δ, the gaps each having an angular width $\xi\frac{\pi}{2}$.

$$D(\xi) = \frac{\xi}{2}\pi \frac{\frac{\sin\theta_2}{1 - \cos\theta_2} + \frac{\sin(\theta_2 + \theta_3)}{1 - \cos(\theta_2 + \theta_3)} - \frac{\sin(\theta_1 + \theta_2)}{1 - \cos(\theta_1 + \theta_2)} - \frac{\sin(\theta_1 + \theta_2 + \theta_3)}{1 - \cos(\theta_1 + \theta_2 + \theta_3)}}{\log_\varepsilon \left\{ \frac{[1 - \cos(\theta_1 + \theta_2)][1 - \cos(\theta_2 + \theta_3)]}{[1 - \cos\theta_2][1 - \cos(\theta_1 + \theta_2 + \theta_3)]} \right\}} + \begin{array}{l}\text{terms in } \xi^2 \\ \text{and higher powers}\end{array} \quad (63)$$

(4.1.1) Some Special Cases of these Results.

If we set

$$\left.\begin{array}{l} \theta_1 = \theta \\ \theta_2 = \pi - \theta \\ \theta_3 = \pi - \theta \end{array}\right\} \quad . \quad . \quad . \quad . \quad (64)$$

we obtain the system of Fig. 3, analysed in Section 2. Substituting conditions (64) in eqn. (61) we obtain

$$C_1(\theta) = \frac{1}{8\pi^2} \log_\varepsilon \frac{2(1 - \cos 2\theta)}{(1 - \cos\theta)(1 + \cos\theta)} \quad . \quad . \quad (65)$$

$$= \frac{1}{4\pi^2} \log_\varepsilon 2 \quad . \quad . \quad . \quad . \quad . \quad (66)$$

as before.

It is of particular interest in this case to investigate the effect, on the mean capacitance per unit length, of small deviations of all the angles from their nominal values [conditions (64)]. Substituting the conditions

$$\left.\begin{array}{ll} \theta_1 = \theta + \delta_1, & \delta_1 \ll \pi/2 \\ \theta_2 = \pi - \theta + \delta_2, & \delta_2 \ll \pi/2 \\ \theta_3 = \pi - \theta + \delta_3, & \delta_3 \ll \pi/2 \end{array}\right\} \quad . \quad . \quad . \quad (67)$$

in expression (62) for the mean capacitance per unit length, and expanding in powers of δ, we find, after some reduction,

$$C_M(\theta) = \frac{1}{4\pi^2} \log_\varepsilon 2 + \frac{1}{16\pi^2} \left\{ \frac{1}{1 - \cos\theta} \left[\frac{\delta_1^2}{2} + \frac{(\delta_1 + \delta_2 + \delta_3)^2}{2} \right] \right.$$

$$\left. + \frac{1}{1 + \cos\theta} \frac{\delta_2^2 + \delta_3^2}{2} + \frac{(\delta_2 + \delta_3)^2}{1 - \cos 2\theta} - \frac{(\delta_1 + \delta_2)^2}{2} \right\}$$

$$+ \text{terms in higher powers of } \delta \quad . \quad . \quad . \quad . \quad (68)$$

which shows that the mean capacitance per unit length remains constant to the first order. It is clear that this result furnishes a corollary to the main theorem given in Section 2, which may be stated (referring to Fig. 1) as follows:

Corollary.—The arithmetic mean of the two direct capacitances per unit length, i.e. $\frac{1}{2}[C(\alpha\beta \text{ to } \gamma\delta) + C(\beta\gamma \text{ to } \alpha\delta)]$, remains constant and equal to C_0, to the first order, for small but arbitrary departures from symmetry.
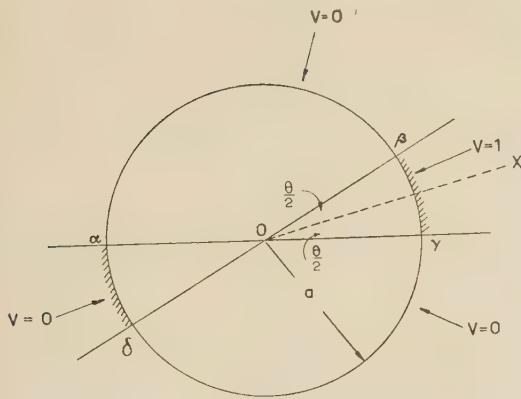
Fig. 7.—Cross-section and co-ordinate system for cylindrical capacitor of circular cross-section (special case).

Another case of interest, shown systematically in Fig. 7, is obtained by setting

$$\left.\begin{array}{l} \theta_1 = \theta \\ \theta_2 = \pi - \theta \\ \theta_3 = \theta \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad (69)$$

When we substitute conditions (69) into eqns. (60) and (61) we obtain

$$C_1(\theta) = \frac{1}{4\pi^2} \log_\varepsilon \frac{2}{1 + \cos\theta} \quad . \quad . \quad . \quad (70a)$$

$$C_2(\theta) = \frac{1}{4\pi^2} \log_\varepsilon \frac{2}{1 - \cos\theta} \quad . \quad . \quad . \quad (70b)$$

The capacitances $C_1(\theta)$ and $C_2(\theta)$ are connected by the identity

$$\varepsilon^{-4\pi^2 C_1(\theta)} + \varepsilon^{-4\pi^2 C_2(\theta)} = 1 \quad . \quad . \quad . \quad . \quad (71)$$

which is analogous to eqn. (32) for the capacitor of rectangular cross-section.

Finally we note that, in general, because of the first-order gap effect [see eqn. (63)], cylindrical capacitors of circular cross-section are not particularly suitable for calculable standards of high precision.

(4.2) Infinite-Plane Cylinder

The system of the infinite-plane cylinder, shown in Fig. 8, arises from application of the Schwarz–Christoffel transformation
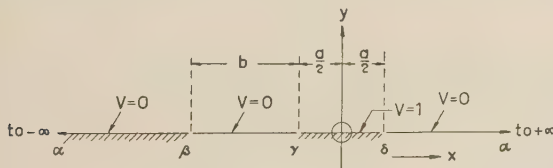
Fig. 8.—Cross-section and co-ordinate system for 'infinite-plane' cylindrical capacitor.

(see Reference 7, p. 197) to a cylindrical capacitor whose cross-section is an arbitrary quadrilateral. In particular, if one diagonal of this quadrilateral is an axis of symmetry, then $a = b$.

Restricting our attention to the field on *one* side of the infinite plane, and choosing co-ordinates as shown in Fig. 8, we have:

Potential at point (x, y).

$$V(x, y) = \frac{2}{\pi} \int_0^\infty \frac{\sin\frac{ka}{2}}{k} \varepsilon^{-ky} \cos kx \, dk, \quad y \geqslant 0 \quad . \quad (72)$$

Surface change density per unit length.

$$\sigma(x) = -\frac{1}{2\pi^2} \int_0^\infty \sin\frac{ka}{2} \cos kx \, dk \quad . \quad . \quad (73)$$

Capacitance per unit length between hatched portions.

$$C(a/b) = \frac{1}{2\pi^2} \int_0^\infty \frac{\sin\frac{ka}{2} \sin k(b + a/2)}{k} \, dk \quad . \quad (74)*$$

$$= \frac{1}{4\pi^2} \log_\varepsilon \left(1 + \frac{a}{b}\right) \quad . \quad . \quad . \quad . \quad (75)$$

Note that $C(1) = \frac{1}{4\pi^2} \log_\varepsilon 2$, which is in agreement with the results of the general theorem.

Mean capacitance per unit length.

$$C_M(a/b) = \tfrac{1}{2}[C(a/b) + C(b/a)] = \frac{1}{4\pi^2} \log_\varepsilon \frac{a + b}{\sqrt{(ab)}} \quad (76)$$

In particular, if $a/b = 1 + \delta$, where $\delta \ll 1$,

$$C_M(1 + \delta) = \frac{1}{4\pi^2} \log_\varepsilon 2 - \frac{\delta^2}{32\pi^2} + \text{terms in higher} \quad \text{powers of } \delta \quad (77)$$

Fraction $D(\xi)$ of total charge on $\alpha\beta$ which lies in a gap of width ξb at β.

$$D(\xi) = \xi \frac{a/b}{(1 + a/b) \log_\varepsilon (1 + a/b)} + \text{terms in higher} \quad \text{powers of } \xi \quad (78)$$

which is of the first order in ξ.
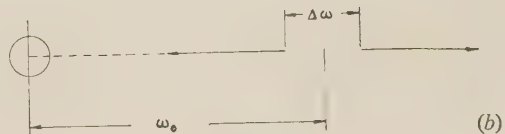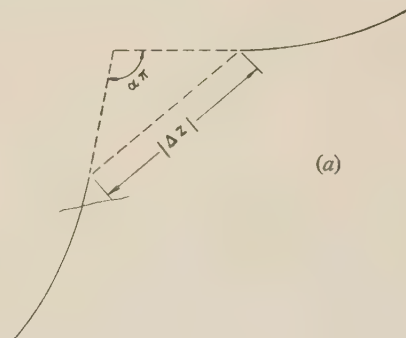
Fig. 9.—The neighbourhood of an insulating gap in a cylindrical capacitor.

(a) Cross-section.
(b) Conformal transformation.

---

* The integral of eqn. (92) is readily put in the form of a Frullani integral (see Reference 12, p. 479), or may be evaluated as a Fourier integral.[13]

It should be noted that the results of this Section on the infinite-plane capacitor provide an alternative proof for the general theorem of Section 1. Thus, if a polygon with $4n$ sides is inscribed in the cross-section of Fig. 1, the result of applying the Schwarz–Christoffel transformation is, in the limit $(n \to \infty)$, the system of Fig. 8 with $a = b$. The proofs of the theorem and corollary then follow from eqns. (75) and (77).

With this procedure in mind, a result concerning the order of magnitude of the error produced by small but finite gaps can be established for the general case.

Thus, considering the neighbourhood of the insulating gap shown in Fig. 1, it can be supposed that any discontinuity in direction of the curve S lies within the gap itself as shown in Fig. 9(a). Application of the Schwarz–Christoffel transformation results in the system shown in Fig. 9(b).

In this neighbourhood, this transformation has the functional form (see Reference 7, p. 197)

$$z = \int F(\omega)(\omega - \omega_0)^{\alpha-1} d\omega + C \quad . \quad . \quad . \quad (79)$$

where C is a constant and $F(\omega)$ is a relatively slowly varying function of ω.

In particular we have

$$\Delta z = \int_{\omega_0 - \Delta\omega/2}^{\omega_0 + \Delta\omega/2} F(\omega)(\omega - \omega_0)^{\alpha-1} d\omega \quad . \quad . \quad . \quad . \quad . \quad (80)$$

$$= F(\omega_0)\left(\frac{\Delta\omega}{2}\right)^{\alpha}\frac{[1-(-1)^{\alpha}]}{\alpha} + \frac{F'(\omega_0)}{1!}\left(\frac{\Delta\omega}{2}\right)^{\alpha+1}\frac{[1-(-1)^{\alpha+1}]}{\alpha+1}$$

$$+ \frac{F''(\omega_0)}{2!}\left(\frac{\Delta\omega}{2}\right)^{\alpha+2}\frac{[1-(-1)^{\alpha+2}]}{\alpha+2} + \text{higher terms} . \quad . \quad (81)$$

i.e.
$$\left[\frac{\Delta z}{F(\omega_0)}\right]^{1/\alpha} = \frac{\Delta\omega}{2}\left\{\frac{[1-(-1)^{\alpha}]}{\alpha}\right.$$

$$\left. + \frac{F'(\omega_0)}{F(\omega_0)}\frac{(\Delta\omega/2)}{1!}\frac{[1-(-1)^{\alpha+1}]}{\alpha+1} + \text{higher terms}\right\}^{1/\alpha} \quad (82)$$

$$= \Delta\omega[a_0 + a_1\Delta\omega + a_2(\Delta\omega)^2 + \ldots] . \quad . \quad . \quad (83)$$

Reversion of this power series (see Reference 10, p. 129) then gives

$$\Delta\omega = b_1\left[\frac{\Delta z}{F(\omega_0)}\right]^{1/\alpha} + b_2\left[\frac{\Delta z}{F(\omega_0)}\right]^{2/\alpha} + b_3\left[\frac{\Delta z}{F(\omega_0)}\right]^{3/\alpha} + \ldots (84)$$

where the b's are functions of the a's of eqn. (83) and hence expressible in terms of $F(\omega)$, its derivatives and α. Taking the modulus of both sides of eqn. (84) we obtain

$$\Delta\omega \leqslant C|\Delta z|^{1/\alpha} + \text{terms in higher powers of } |\Delta z| . \quad (85)$$

Eqn. (78) shows that the fraction of charge in a narrow gap in the infinite plane system is linear to the first order in the gap width. Hence, from eqn. (85), it can be seen that a gap of width $|\Delta z|$ in the system shown in Fig. 1 will produce an error in the direct capacitance whose magnitude is proportional, to the first order, to $|\Delta z|^{1/\alpha}$, where $\alpha\pi$ is the angle between tangents to the electrode surfaces at the gap.

For example, for the rectangular system treated in Section 3, we have $\alpha = \frac{1}{2}$ at every gap, and hence the fractional error in capacitance is proportional to (gap width)^2, which is in agreement with eqn. (57).

It is clear from eqn. (85) that a configuration highly insensitive to the width of the gaps used would result from placing the gap in a re-entrant position such that α is as small as possible. A configuration which achieves this objective, and at the same time has considerable advantages in practice, is shown in Fig. 10.
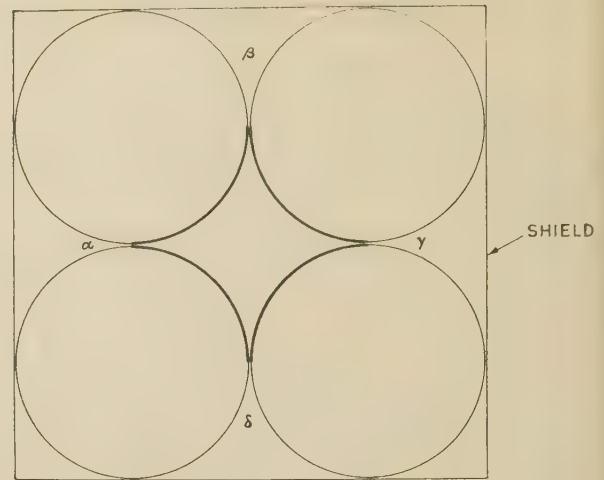
Fig. 10.—Cross-section of a practical cylindrical capacitor.

Calculable capacitors based on this design have been constructed, and will be described in detail in a forthcoming publication, by the author's colleague A. M. Thompson, to whom this configuration is due.

Of some practical interest is the effect, on the direct capacitance of a cylindrical capacitor, of filling part of the interior of the cylinder with material whose dielectric constant differs from unity. To obtain some idea of the magnitude of this effect, particularly in the limiting case of thin dielectric films on the electrode surfaces (representing a possible surface contamina-
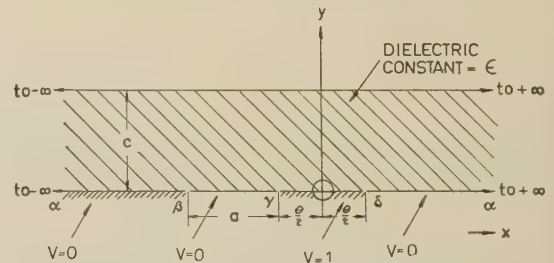
Fig. 11.—Cross-section of 'infinite-plane' cylindrical capacitor with dielectric film.

tion), an analysis of the symmetrical infinite-plane system shown in Fig. 11 has been made, with the following results.

Potential at point (x, y) in Region 1, $y \leqslant c$.

$$V_1(x,y) = \frac{2}{\pi}\int_0^\infty \sin\frac{ka}{2}\cos kx\left[\frac{\left(1+\frac{1}{\epsilon}\right)\varepsilon^{-ky} + \left(1-\frac{1}{\epsilon}\right)\varepsilon^{ky}\varepsilon^{-2kc}}{\left(1+\frac{1}{\epsilon}\right) + \left(1-\frac{1}{\epsilon}\right)\varepsilon^{-2kc}}\right]\frac{dk}{k}$$

$$. \quad . \quad . \quad (86)$$

Potential at point (x, y) in Region 2, $y \geqslant c$.

$$V_2(x,y) = \frac{4}{\pi}\int_0^\infty \frac{\sin\frac{ka}{2}\cos kx\,\varepsilon^{-ky}}{\left(1+\frac{1}{\epsilon}\right) + \left(1-\frac{1}{\epsilon}\right)\varepsilon^{-2kc}}\frac{dk}{k} . \quad (87)$$

Surface charge density per unit length.

$$\sigma(x) = -\frac{\epsilon}{2\pi^2} \int_0^\infty \sin\frac{ka}{2} \cos kx \frac{\left(1+\frac{1}{\epsilon}\right)-\left(1-\frac{1}{\epsilon}\right)\varepsilon^{-2kc}}{\left(1+\frac{1}{\epsilon}\right)+\left(1-\frac{1}{\epsilon}\right)\varepsilon^{-2kc}} \, dk \quad (88)$$

Direct capacitance per unit length between hatched portions.

$$C\left(\epsilon;\frac{c}{a}\right) = \frac{1}{4\pi^2}\log_\varepsilon 2$$
$$-\frac{\epsilon}{4\pi^2}\sum_{s=1}^{\infty}\left(\frac{1-\frac{1}{\epsilon}}{1+\frac{1}{\epsilon}}\right)^s (-1)^s \log_\varepsilon\left[\frac{1+4\left(s\frac{c}{a}\right)^2}{1+\left(s\frac{c}{a}\right)^2}\right]$$
$$\cdots \quad (89)$$

In particular, for $c/a \ll 1$, corresponding to thin films, we find:

$$C\left(\epsilon;\frac{c}{a}\right) = \frac{1}{4\pi^2}\log_\varepsilon 2 + \frac{3}{16\pi^2}\left(\frac{c}{a}\right)^2\left(1-\frac{1}{\epsilon^2}\right) + \text{terms in} \left(\frac{c}{a}\right)^4 \text{and higher} \quad (90)$$

which shows that thin dielectric films produce only a second-order effect on the capacitance.

We note the following special cases of eqn. (89):

$$C\left(1;\frac{c}{a}\right) = C(\epsilon;0) = \frac{1}{4\pi^2}\log_\varepsilon 2 \quad \cdots \quad (91)$$

$$C(\epsilon;\infty) = \frac{\epsilon}{4\pi^2}\log_\varepsilon 2 \quad \cdots \quad (92)$$
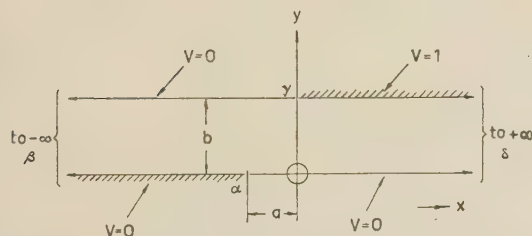
Fig. 12.—Cross-section and co-ordinate system for 'double infinite-plane' cylindrical capacitor.

(4.3) Double Infinite-Plane Cylinder

With the co-ordinates and origin as shown in Fig. 12, we have:

Potential at a point (x, y), where $0 \leqslant y \leqslant b$.

$$V(x,y) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left[\frac{1}{jk}+\pi\delta(k)\right]\varepsilon^{jkx}\frac{\sinh ky}{\sinh kb}\,dk \quad (93)$$

where $\delta(k)$ is the unit impulse function.

Surface charge density per unit length.

$$\sigma(x) = \frac{1}{8\pi^2}\int_{-\infty}^{+\infty}\left[\frac{1}{jk}+\pi\delta(k)\right]\frac{k\varepsilon^{jkx}}{\sinh kb}\,dk \quad (94)$$

$$= \frac{1}{4\pi b}\frac{\varepsilon^{\pi x/b}}{1+\varepsilon^{\pi x/b}} \quad \cdots \quad (95)$$

The integral occurring in eqn. (94) is readily evaluated by contour integration, using an infinite semicircle in the lower half-plane and an indentation of the real axis at the origin.

Capacitance per unit length between hatched portions.

$$C_1\left(\frac{a}{b}\right) = \frac{1}{4\pi^2}\log_\varepsilon(1+\varepsilon^{-\pi a/b}) \quad \cdots \quad (96)$$

When one gap is directly above the other, we have, on setting $a = 0$,

$$C_1(0) = \frac{1}{4\pi^2}\log_\varepsilon 2 \quad \cdots \quad (97)$$

which agrees with the general theorem.

Capacitance per unit length between unhatched portions.

$$C_2\left(\frac{a}{b}\right) = \frac{a}{4\pi b}+\frac{1}{4\pi^2}\log_\varepsilon(1+\varepsilon^{-\pi a/b}) \quad \cdots \quad (98)$$

Note particularly that

$$C_2\left(\frac{a}{b}\right) - C_1\left(\frac{a}{b}\right) = \frac{a}{4\pi b} \quad \cdots \quad (99)$$

which is linear in a/b.

Mean capacitance per unit length.

$$C_M\left(\frac{a}{b}\right) = \frac{a}{8\pi b}+\frac{1}{4\pi^2}\log_\varepsilon(1+\varepsilon^{-\pi a/b}) \quad \cdots \quad (100)$$

and if we put $a/b = \delta$, where $\delta \ll 1$, and expand in powers of δ we find

$$C_M(\delta) = \frac{1}{4\pi^2}\log_\varepsilon 2 + \frac{\delta^2}{16} + \text{terms in } \delta^3 \text{ and higher} \quad (101)$$

which is in agreement with the corollary to the general theorem.

Fraction $D(\xi)$ of total charge on $\alpha\beta$ which lies in a gap at α whose width is a fraction ξ of the interplane spacing b.

$$D(\xi) = \xi\frac{\pi\varepsilon^{-\pi a/b}}{(1+\varepsilon^{-\pi a/b})\log_\varepsilon(1+\varepsilon^{-\pi a/b})} + \text{terms in } \xi^2 \text{ and higher} \quad (102)$$

which result agrees with the conclusions of eqn. (85).

(5) CONCLUSIONS

A theorem has been established which states that the direct capacitance per unit length of a certain class of cylindrical 3-terminal capacitors is a constant.

A first essential requirement in this class of capacitors is that the cylinder cross-section shall have at least one axis of symmetry, but it has been shown that when the symmetry is slightly disturbed in an arbitrary way, the residual error in the mean of the two cross-capacitances (per unit length) is of the second order.

Finally, the effect of the necessarily finite insulating gaps has been examined, in an approximate way, and it has been shown that, providing that the gaps are placed in suitably re-entrant corners, their effect may be made negligible.

These investigations show that cylindrical 3-terminal capacitors are extremely attractive as calculable capacitance standards, one precision length measurement only being required to compute the capacitance.

(6) ACKNOWLEDGMENTS

during the course of this work. He also wishes to thank Mr. D. A. Mustard for helpful discussions concerning some of the theoretical work presented. Finally he would like to thank the referees whose comments have been of considerable assistance during the revision of this paper.

(7) REFERENCES

(1) ROSA, E. B., and DORSEY, N. E.: 'A New Determination of the Ratio of the Electromagnetic to the Electrostatic Unit of Electricity', *Bulletin of Bureau of Standards*, 1907, **3**, p. 433.

(2) MOON, C., and SPARKS, M. C.: 'Standards for Low Values of Direct Capacitance', *National Bureau of Standards Journal of Research*, 1948, **41**, p. 497.

(3) SNOW, C.: 'A Standard of Small Capacitance', *ibid.*, 1947, **42**, p. 848.

(4) SLONCZEWSKI, T.: 'Absolute Capacitance Standard with Resistive Shield', *Review of Scientific Instruments*, 1947, p. 848.

(5) SNOW, C.: 'Formulas for computing Capacitance and Inductance', *National Bureau of Standards Circular No. 544*.

(6) THOMPSON, A. M., and LAMPARD, D. G.: 'A New Theorem in Electrostatics with Applications to Calculable Standards of Capacitance', *Nature*, 1956, **177**, p. 888.

(7) COPSON, E. T.: 'Theory of Functions of a Complex Variable' (Oxford University Press, 1948).

(8) CAMPBELL, G. A.: 'Direct Capacitance Measurement', *Bell System Technical Journal*, 1922, **1**, p. 18.

(9) SMYTHE, W. R.: 'Static and Dynamic Electricity' (McGraw-Hill, 1939).

(10) WHITTAKER, E. T., and WATSON, G. N.: 'A Course of Modern Analysis' (Cambridge University Press, 1950).

(11) WOODS, L. C.: 'Subsonic Plane Flow in an Annulus or a Channel with Spacewise Periodic Boundary Conditions', *Proceedings of the Royal Society*, 1955, **229**, p. 63–85.

(12) BROMWICH, T. J. I'A.: 'An Introduction to the Theory of Infinite Series' (Macmillan and Co., 1949).

(13) CAMPBELL, G. A., and FOSTER, R. M.: 'Fourier Integrals for Practical Application—Pair No. 644.4' (D. Van Nostrand, 1951).

(8) APPENDIX

Evaluation of $\sum_{n=0}^{\infty} \dfrac{1}{(2n+1)\sinh\left[(2n+1)\pi\beta\right]}$

By using the result

$$\frac{1}{\sinh\theta} = 2\sum_{s=0}^{\infty} \varepsilon^{-(2s+1)\theta} \quad . \quad . \quad . \quad (103)$$

we may write

$$\sum_{n=0}^{\infty} \frac{1}{(2n+1)\sinh\left[(2n+1)\pi\beta\right]}$$

$$= 2\sum_{s=0}^{\infty}\left[\sum_{n=0}^{\infty}\frac{1}{(2n+1)}\varepsilon^{-(2s+1)(2n+1)\pi\beta}\right] \quad . \quad (104)$$

$$= 2\sum_{s=0}^{\infty} -\pi(2s+1)\sum_{n=0}^{\infty}\int_{\infty}^{\beta}\varepsilon^{-(2n+1)(2s+1)\pi\beta}d\beta \quad (105)$$

$$= \sum_{s=0}^{\infty} -\pi(2s+1)\int_{\infty}^{\beta}\frac{d\beta}{\sinh\left[(2s+1)\pi\beta\right]} \quad . \quad (106)$$

$$= -\sum_{s=0}^{\infty}\log_{\varepsilon}\left\{\tanh\left[(2s+1)\frac{\pi\beta}{2}\right]\right\} \quad . \quad . \quad (107)$$

$$= \log_{\varepsilon}\prod_{s=0}^{\infty}\coth\left[(2s+1)\frac{\pi\beta}{2}\right] \quad . \quad . \quad (108)$$

$$= \tfrac{1}{2}\log_{\varepsilon}\prod_{s=0}^{\infty}\left[\frac{1+\varepsilon^{-\pi\beta(2s+1)}}{1-\varepsilon^{-\pi\beta(2s+1)}}\right] \quad . \quad . \quad . \quad (109)$$

where, in passing from eqn. (106) to eqn. (107), we have used the result

$$\int\frac{dx}{\sinh ax} = \frac{1}{a}\log_{\varepsilon}\left(\tanh\frac{ax}{2}\right) \quad . \quad . \quad (110)$$

Now the theta functions $\theta_{3,4}(z;q)$ have the well-known infinite product representation (see Reference 10, p. 469):

$$\theta_3(z;q) = G\prod_{n=0}^{\infty}(1+2q^{2n+1}\cos 2z + q^{4n+2}) \quad . \quad (111a)$$

$$\theta_4(z;q) = G\prod_{n=0}^{\infty}(1-2q^{2n+1}\cos 2z + q^{4n+2}) \quad . \quad (111b)$$

where

$$G = \prod_{n=1}^{\infty}(1-q^{2n}) \quad . \quad . \quad . \quad (111c)$$

and thus it follows that

$$\frac{\theta_3(0;q)}{\theta_4(0;q)} = \prod_{n=0}^{\infty}\left(\frac{1+q^{2n+1}}{1-q^{2n+1}}\right)^2 \quad . \quad . \quad (112)$$

When we set $q = \varepsilon^{-\pi\beta}$ and make use of eqns. (112) and (109), we obtain

$$\sum_{n=0}^{\infty}\frac{1}{(2n+1)\sinh(2n+1)\pi\beta} = \frac{1}{2}\log_{\varepsilon}\frac{\theta_3(0;\varepsilon^{-\pi\beta})}{\theta_4(0;\varepsilon^{-\pi\beta})} \quad (113)$$

which is the required result.

TORQUE-ANGLE ANALYSIS OF THE SCHRAGE MOTOR

By N. KESAVAMURTHY, M.A., M.Sc.Tech., Graduate, and R. E. BEDFORD, B.Sc., Ph.D.

SUMMARY

Reliable equations of performance for the Schrage motor are deduced in terms of the angle between the stator and the rotor magnetomotive force waves, resulting in an original treatment. The expressions for torque, slip and current in terms of this angle are derived and are discussed with the aid of graphs and locus diagrams.

LIST OF PRINCIPAL SYMBOLS

The primary, secondary and tertiary windings will be referred to as windings 1, 2 and 3, respectively.

V_1 = Applied voltage per phase to winding 1.

f = Frequency of supply voltage.

r = Ratio of effective turns between brushes of winding 3 to effective turns of winding 2.

ρ = Electrical angle by which brush axis leads axis of corresponding phase of winding 2.

k = Ratio of effective turns per phase of winding 1 to those of winding 2.

s = Slip.

R_1, R_2 and R_3 = Effective resistances per phase of windings 1, 2 and 3, respectively.

X_{11}, X_{21} and X_{31} = Leakage reactances per phase at frequency f, of windings 1, 2 and 3, respectively.

X_{11}, X_{22} and X_{33} = Self-reactances per phase (exclusive of the leakage reactances) of windings 1, 2 and 3, respectively; all quantities being referred to windings 1 at frequency f. Each of these reactances is due to the fluxes produced by all three phases of the particular winding.

$X_{32}, X_{12}, \ldots$ = Mutual-reactances per phase between the windings indicated by the subscripts, all quantities being referred to winding 1 at frequency f.

δ = Space angle by which magnetomotive-force wave of winding 1 *leads* that of winding 2, electrical degrees.

I_1 and I_2 = Phase currents of windings 1 and 2, respectively.

$p = I_2/I_1$.

(1) INTRODUCTION

The principle of the Schrage motor is well known. Various methods of analysis are given in the literature.[1,2,3,4,5,6] All these analyses rest on certain simplifying assumptions; the work of Gibbs,[7] using Kron's[8] tensor analysis, being the most rigorous and accurate.

In the paper an attempt is made to arrive at the equations of performance of the machine based on the angle between the magnetomotive force (m.m.f.) waves F_1 and F_2 of the respective windings 1 and 2. This results in an original treatment and gives a greater insight into the physical phenomena involved in the performance of the machine. Besides, these equations are accurate, and, in fact, the equations of Gibbs can be deduced therefrom on making suitable simplifying assumptions.

(2) GENERAL

(2.1) Description of the Machine

The Schrage motor is fundamentally an induction machine; the primary winding (winding 1) is on the rotor and the secondary winding (winding 2) is on the stator and receives an injected e.m.f. from the brushes riding on the commutator of a d.c. winding (winding 3) on the rotor. The d.c. winding acts as a frequency-changer, changing the frequency from that of the supply to the (slip) frequency of the induced e.m.f. in winding 2. In addition, the magnitude of the injected voltage can be controlled by varying the angle of separation of brushes, thus altering the ratio of transformation. Furthermore, the brush system as a whole can be shifted from the position in which the induced voltage in winding 2 and the injected voltage supplied by winding 3 are in direct phase opposition. This constitutes the brush *shift*.

(2.2) The Method of Analysis

For balanced steady-state operation, balanced polyphase currents flow in all the windings; those in windings 1 and 3 are at the supply frequency and that in winding 2 is at the slip frequency. These sets of currents flowing in windings 1, 2 and 3 give rise to revolving m.m.f. waves F_1, F_2 and F_3, respectively, these being relatively at rest and rotating at slip frequency with reference to the stator. The angle between m.m.f. waves F_2 and F_3 is equal to the angle of brush shift, ρ, and is independent of the load. However, the angle between F_1 and F_2 depends on the torque, and this angle of separation, δ, will be termed the *torque angle*. The magnitudes of F_1, F_2 and F_3 are proportional to their respective currents; that of F_3 being, in addition, dependent on the brush separation.

(2.3) Simplifying Assumptions

The following assumptions are made in the development of the theory:

(*a*) Saturation is neglected; this is particularly valid in view of the fact that usually the machine is designed for low flux densities. Hence for normal operation all reactances may be considered to be constant.

(*b*) Only the fundamentals of the m.m.f. waves are considered, all other harmonics being ignored as secondary effects.

(*c*) Iron losses are neglected, since they depend on the resultant flux density and frequency; for the stator they are practically constant, whilst for the rotor they are functions of the slip. Consequently, any analysis which attempts the inclusion of these losses becomes involved. In the paper these losses are initially ignored, although an approximation can later be made as in the case of induction motors.

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Kesavamurthy and Dr. Bedford are in the Electrical Engineering Department, Indian Institute of Technology, Kharagpur.

(d) Stray load losses are neglected.

(e) The influence of the commutating coil on the m.m.f.'s and fluxes of the machine is ignored.

(f) In the formulation of the equations of performance all resistances are assumed to be constant; this is in error particularly with reference to R_3, because of the brush-contact resistance, which is a function of the current. However, in actual computation the correct procedure would be to calculate a sufficient number of secondary currents from the equations using the appropriate value of R_3, this being obtained by successive approximations based on the experimental curve relating R_3 to its currents.

(3) EQUATIONS OF PERFORMANCE

(3.1) M.M.F. Waves F_1, F_2 and F_3

For a constant shaft torque and for a fixed setting of the brushes (r and ρ) the machine runs at a steady speed corresponding to the slip s. The currents in windings 1 and 3 are at the supply frequency f, and that in winding 2 is at the slip frequency sf. However, each of the m.m.f.'s rotates at synchronous speed with respect to the rotor conductors and hence is relatively at rest. The magnitudes of these m.m.f.'s are proportional to I_1T_1, I_2T_2 and rI_2T_3, respectively, where r is governed by the
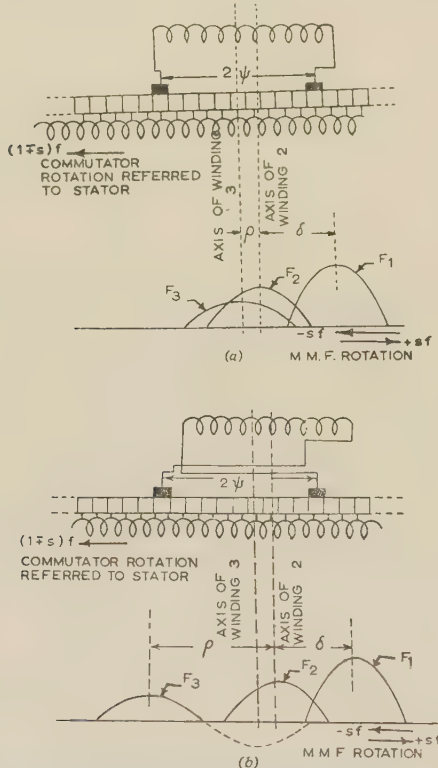
Fig. 1.—M.M.F. relationships.

(a) For $\rho < 180°$.
(b) For $\rho > 180°$.

brush separation. With a uniform distribution of the commutator winding, if r_{max} is the value of r for a brush separation of 180 electrical degrees, then for a brush separation of 2ψ, the value of r is given by the relation

$$r = r_{max} \sin \psi$$

Consider, next, the distribution in space of these m.m.f. waves. As seen from Figs. 1(a) and 1(b), F_2 lags behind F_1 by the torque

angle δ, whilst F_2 and F_3 are separated by the angle ρ of the brush shift. Obviously ρ can assume any value from 0 to 360 electrical degrees depending on the position of the brushes; in particular, for values of ρ between 180° and 360° the brushes cross over as shown in Fig. 1(b).

(3.2) Fluxes and Reactances

The distribution of the fluxes for assumed currents in all the windings is shown schematically in Fig. 2. These are as follows:

(a) The leakage fluxes Φ_{1l}, Φ_{2l} and Φ_{3l} of the windings 1, 2 and 3, respectively, linking with their respective currents.
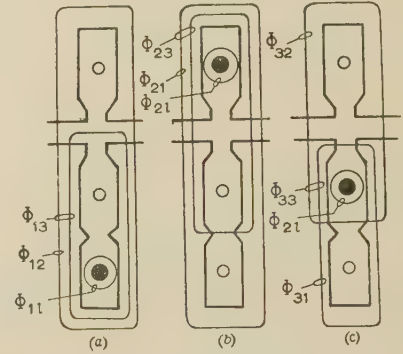
Fig. 2.—Fluxes due to currents in windings.

(a) Fluxes due to currents in winding 1.
(b) Fluxes due to currents in winding 2.
(c) Fluxes due to currents in winding 3.

(b) The self-fluxes Φ_{11}, Φ_{22} and Φ_{33}, which are the total fluxes (exclusive of leakage fluxes) due to currents in the windings 1, 2 and 3, respectively.

(c) The mutual fluxes Φ_{13}, Φ_{12}, Φ_{23}, Φ_{31}, Φ_{21} and Φ_{32} mutual to the windings indicated by the subscripts, the first subscript denoting the actuating winding.

The leakage reactances X_{1l}, X_{2l} and X_{3l} correspond to the leakage fluxes Φ_{1l}, Φ_{2l} and Φ_{3l}, respectively; the magnetizing reactances X_{11}, X_{22}, X_{33} correspond to Φ_{11}, Φ_{22} and Φ_{33}, respectively; and the reactances X_{12}, X_{23}, etc., correspond, respectively, to the fluxes Φ_{12}, Φ_{23}, etc. All these reactances are reckoned at the supply frequency f, and are referred to the turns of winding 1.

(3.3) M.M.F. Relationships

To formulate the equations of performance for voltages it is essential to determine the rate of change of flux linkages associated with each winding. This in turn depends on the action of each of the m.m.f. waves on the particular winding under consideration. At this stage it becomes necessary to postulate a hypothetical current in each winding called the *exciting current*, and defined as that current which, by its independent action in its winding, can produce the total flux linkages (exclusive of the self-leakage flux) associated with this winding and caused by all the m.m.f.'s. Denoting these exciting currents in windings 1 and 2 by I_{01} and I_{02}, respectively, the balancing equations for the m.m.f.'s are as follows:

For the stator

$$\check{I}_{01}T_1 = k_1I_1T_1\varepsilon^{j\alpha} + \frac{X_{12}}{X_{11}}k_1I_2T_2\varepsilon^{j(\alpha-\delta)} + k_1rI_2T_2\varepsilon^{j(\alpha-\delta-\rho)} \quad (1)$$

For the rotor at sub-synchronous speeds

$$\check{I}_{02}T_2 = k_2\frac{X_{12}}{X_{22}}I_1T_1\varepsilon^{j(\beta+\delta)} + k_2I_2T_2\varepsilon^{j\beta} + k_2\frac{X_{32}}{X_{22}}rI_2T_2\varepsilon^{j(\beta-\rho)} \quad (2a)$$

For the rotor at super-synchronous speeds

$$\dot{I}_{02}T_2 = \dot{k}_2 \frac{X_{12}}{X_{22}} I_1 T_1 \varepsilon^{j(\beta-\delta)} + \dot{k}_2 I_2 T_2 \varepsilon^{j\beta} + \dot{k}_2 \frac{X_{32}}{X_{22}} r I_2 T_2 \varepsilon^{j(\beta+\rho)} \quad (2b)$$

where $\dot{k}_1$ and $\dot{k}_2$ are the revolving unit vectors at frequencies f and sf, respectively, and α and β are the angles by which the currents I_1 and I_2, respectively, lead $\dot{k}_1$ and $\dot{k}_2$. In the formulation of these equations it must be borne in mind that the various n.m.f.'s act on different magnetic circuits, as is clear from Fig. 2. Consequently, the equivalent ampere-turns required in, say, winding 3 to produce the same flux linkages with itself as those caused by the ampere-turns of winding 2 will be given by the relation:

(ampere-turns of winding 3)

$$= \text{(ampere-turns of winding 2)} \frac{\text{reluctance for } \Phi_{33}}{\text{reluctance for } \Phi_{23}},$$

$$= \text{(ampere-turns of winding 2)} \frac{X_{23}}{X_{33}}$$

This accounts for the presence of the reactance ratios in the above equations. The various steps leading to these equations are given in Sections 9.1 and 9.2.

(3.4) Equations of Voltages

Each of the exciting currents can be split up into active and reactive components with reference to the e.m.f.'s induced in the windings. Thus for winding 1:

$$I_{r1} = I_{01} \cos \sigma \times \varepsilon^{-j\sigma}$$

$$I_{a1} = I_{01} \sin \sigma \times \varepsilon^{j(\pi/2-\sigma)}$$

Where the suffixes r and a denote the reactive and active components, respectively, and σ is the angle by which I_{r1} lags behind I_{b1}. For the assumption of negligible iron losses, the induced electromotive forces V'_1, V'_2 and V'_3 in the respective windings 1, 2 and 3 are given by the following relations:

For the stator

$$\dot{V}'_1 = j\dot{I}_1 \left(X_{11} + \varepsilon^{-j\delta} \frac{p}{k} X_{12} + rX_{13}\varepsilon^{-j\rho} \right) \quad . \quad . \quad (3)$$

For the rotor at sub-synchronous speeds

$$\dot{V}'_2 = j\dot{I}_2 \left(\frac{sX_{12}}{k^2} \frac{k}{p} \varepsilon^{j\delta} + \frac{sX_{22}}{k^2} + \frac{sX_{32}}{k^2} r\varepsilon^{-j\rho} \right) \quad . \quad (4a)$$

For the rotor at super-synchronous speeds

$$\dot{V}'_2 = j\dot{I}_2 \left(\frac{sX_{12}}{k^2} \frac{k}{p} \varepsilon^{-j\delta} + \frac{sX_{22}}{k^2} + \frac{sX_{32}}{k^2} r\varepsilon^{+j\rho} \right) \quad . \quad (4b)$$

For the tertiary at sub-synchronous speeds

$$\dot{V}'_3 = j\dot{I}_2 \left(\frac{rX_{13}}{k^2} \frac{k}{p} \varepsilon^{j(\delta+\rho)} + \frac{rX_{23}}{k^2} \varepsilon^{j\rho} + \frac{r^2}{k^2} X_{33} \right) \quad . \quad (5a)$$

For the tertiary at super-synchronous speeds

$$\dot{V}'_3 = -j\dot{I}_2 \left[\frac{rX_{13}}{k^2} \frac{k}{p} \varepsilon^{-j(\delta+\rho)} + r\left(\frac{X_{23}}{k^2}\right) \varepsilon^{-j\rho} + \frac{r^2}{k^2} X_{33} \right] \quad (5b)$$

The voltage equations are therefore given by

$$\dot{V}_1 = \dot{I}_1(R_1 + jX_{1l}) + \dot{V}'_1$$

$$= \dot{I}_1 \left[R_1 + j(X_{11} + X_{1l}) + j\varepsilon^{-j\delta} \frac{p}{k} (X_{12} + rX_{13}\varepsilon^{-j\rho}) \right] \quad (6)$$

and

$$\dot{V}_2 + \dot{V}_3 \equiv O = \dot{I}_2 \left[\frac{1}{k^2}(R_2 + r^2 R_3) + j\frac{1}{k^2}(sX_{2l} + r^2 sX_{3l}) \right] + \dot{V}'_2 + \dot{V}'_3 \quad (7)$$

since windings 2 and 3 are in series.

(3.5) Equations of Performance

Eqns. (6) and (7) are obtained in terms of the ratio $p = I_2/I_1$ and the torque angle δ. The expressions for k/p can be derived from either of the two sets of eqns. (7), (4a) and (5a) or (7), (4b) and (5b), and it is seen that in either case,

$$\frac{k}{p} = \frac{(R_2 + r^2 R_3) + sX_{23}r \sin \rho - rX_{23} \sin \rho}{sX_{12} \sin \delta + rX_{13} \sin(\delta + \rho)}$$

$$= -\frac{s(X_{2l} + r^2 X_{3l} + X_{22}) + sX_{23}r \cos \rho + rX_{23} \cos \rho + r^2 X_{33}}{sX_{12} \cos \delta + rX_{13} \cos(\delta + \rho)} \quad (8)$$

provided that s is reckoned positive for sub-synchronous and negative for super-synchronous operation.

Hence eqn. (7) could be rewritten as

$$0 = \frac{\dot{I}_2}{k^2} \left\{ R_2 + r^2 R_3 + j\frac{k}{p} \varepsilon^{j\delta}(sX_{12} + rX_{13}\varepsilon^{j\rho}) \right. $$
$$\left. + j[s(X_{2l} + r^2 X_{3l} + X_{22} + X_{23}r\varepsilon^{-j\rho}) + rX_{23}\varepsilon^{j\rho} + r^2 X_{33}] \right\} \quad (9)$$

Eliminating s from relations (8) gives the quadratic in p as

$$\left(\frac{p}{k}\right)^2 n + \frac{p}{k}(h \sin \delta + l \cos \delta) + m = 0 \quad . \quad . \quad (10)$$

where

$$h = \{rX_{13}[rX_{23}(\cos^2 \rho - \sin^2 \rho) + (X_{22} + X_{2l} + r^2 X_{3l}) \cos \rho] - rX_{12}X_{23} \cos \rho - r^2 X_{12}X_{33}\}$$

$$l = [rX_{13}(X_{22} + X_{2l} + r^2 X_{3l}) \sin \rho + rX_{12}X_{23} \sin \rho - X_{12} (R_2 + r^2 R_3) + 2r^2 X_{13}X_{23} \sin \rho \cos \rho]$$

$$m = rX_{13}X_{12} \sin \rho$$

$$n = \{[rX_{23} \sin \rho - (R_2 + r^2 R_3)][rX_{23} \cos \rho + X_{22} + X_{2l} + r^2 X_{3l}] + r^2 X_{23}^2 \sin \rho \cos \rho + r^3 X_{33}X_{23} \sin \rho\}$$

(3.5.1) Equations for Slip.

From eqns (8) the relation in s is given by the quadratic

$$s^2(a \cos \delta + b \sin \delta) + s(c \cos \delta + d \sin \delta) + (f \cos \delta + g \sin \delta) = 0$$

or

$$s = \frac{1}{2} \left\{ -\left(\frac{c \cos \delta + d \sin \delta}{a \cos \delta + b \sin \delta}\right) \pm \sqrt{\left[\left(\frac{c \cos \delta + d \sin \delta}{a \cos \delta + b \sin \delta}\right)^2 - 4\left(\frac{f \cos \delta + g \sin \delta}{a \cos \delta + b \sin \delta}\right)\right]} \right\} \quad (11)$$

$$a = rX_{23}X_{12} \sin \rho$$

$$b = X_{12}(X_{2l} + r^2 X_{3l} + X_{22} + X_{23}r \cos \rho)$$

$$c = [r^2 X_{23}X_{13} \sin \rho \cos \rho + X_{12}(R_2 + r^2 R_3 - rX_{23} \sin \rho) + rX_{13} \sin \rho(X_{2l} + r^2 X_{3l} + X_{22} + X_{23}r \cos \rho)]$$

$$d = [X_{12}(rX_{23} \cos \rho + r^2 X_{33}) + rX_{13} \cos \rho(X_{2l} + r^2 X_{3l} + X_{22} + X_{23}r \cos \rho) - r^2 X_{23}X_{13} \sin^2 \rho]$$

$$f = [rX_{13} \cos \rho (R_2 + r^2R_3 - rX_{23} \sin \rho)$$
$$+ rX_{13} \sin \rho (rX_{23} \cos \rho + r^2X_{33})]$$

$$g = [rX_{13} \cos \rho (rX_{23} \cos \rho + r^2X_{33})$$
$$- rX_{13} \sin \rho (R_2 + r^2R_3 - rX_{23} \sin \rho)]$$

In computation it is sufficient to work with either of the signs before the root, as values corresponding to the other sign are obtainable by replacing δ by $\pi + \delta$. Obviously δ can take only such values as would make the quantity under the root real. The full significance of δ is brought out in Section 9.3.

(3.5.2) Equations of Voltage.

Solving for p from eqn. (9) and substituting in eqn. (6) it is found that

$$\dot{V}_1 = \dot{I}_1 \left(R_1 + j(X_{1l} + X_{11}) + j(X_{12} + rX_{13}\varepsilon^{-j\rho})\varepsilon^{-j\delta} \right.$$

$$\left\{ -\left(\frac{h \sin \delta + l \cos \delta}{2n} \right) \pm \sqrt{\left[\left(\frac{h \sin \delta + l \cos \delta}{2n} \right)^2 - \frac{m}{n} \right]} \right\} \right)$$

$$. \quad . \quad . \quad . \quad (12)$$

As in the case of s, here also it suffices to work with either one of the signs before the root.

Alternatively, by eliminating $k\varepsilon^{j\delta}/p$ from eqns. (6) and (9) the expression for V_1 is related in terms of s only as

$$\dot{V}_1 = \frac{C + jD}{A + jB} \dot{I}_1 \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where

$A = R_2 + r^2R_3 - r \sin \rho \, X_{23}(1 - s)$

$B = s(X_{2l} + X_{22}) + r \cos \rho \, X_{23}(1 + s) + r^2(X_{33} + sX_{3l})$

$C = AR_1 - B(X_{1l} + X_{11}) + sX_{12}^2 + r^2X_{13}^2 + rX_{12}X_{13} \cos \rho(1 + s)$

$D = BR_1 + A(X_{1l} + X_{11}) + rX_{12}X_{13}(1 - s) \sin \rho$

From eqn. (8) it is seen that

$$\left(\frac{k}{p} \right)^2 \equiv k^2 \left(\frac{I_1}{I_2} \right)^2 = \frac{[(R_2 + r^2R_3) + \sin\rho(srX_{23} - rX_{23})]^2 + [rX_{23}\cos\rho(s+1) + s(X_{22} + X_{2l} + r^2X_{3l} + r^2X_{33})]^2}{(sX_{12})^2 + (rX_{23})^2 + (2srX_{12}X_{13} \cos \rho)}$$

whence, knowing I_1, I_2 can be computed.

(3.5.3) Equation for Torque.

The torque per phase, in synchronous watts, is given by

$$T = Rl\dot{V}_1\bar{I}_1 - I_1^2R_1 - I_1^2p^2(R_2 + r^2R_3) \quad . \quad . \quad (14)$$

The foregoing eqns. (8), (11), (12) and (14) are the equations of performance and are sufficient to determine the primary and secondary currents, the slip, the power factor, the input and output, and the efficiency at any value of the parameter, δ, for any specified brush position (including both separation and shift).

The torque obtained is the gross torque in that it includes all those losses usually considered independent of load, such as iron loss, friction and windage loss, and parasitic-current losses under the brushes. A torque equivalent to these steady losses must be calculated and subtracted from the gross torque to obtain the net output torque.

(4) LOCUS DIAGRAMS

For any particular value of ψ and ρ, the equation of the locus of the point representing the primary current I_1 can be established by using either eqn. (12) expressed in terms of δ, or eqn. (13) derived in terms of s.

As shown in Section 9.3 the locus of the impedance function defined by eqn. (12) is a circle for varying values of δ, the parameters of the circle being defined as follows:

$$Centre: \left[R_1 - (X_{12} + rX_{13} \cos \rho)\frac{h}{2n} - \frac{lrX_{13}\sin\rho}{2n} \right.$$

$$\left. + j\left[(X_{11} + X_{1l}) + \frac{hrX_{13}\sin\rho}{2n} - \frac{l}{2n}(X_{12} + rX_{13}\cos\rho) \right] \right.$$

$$= F + jG, \text{ say}$$

$$Radius: R = \frac{1}{2n}[(X_{12} + rX_{13}\cos\rho)^2$$
$$+ (rX_{13}\sin\rho)^2]^{1/2}[l^2 + h^2 - 4mn]^{1/2}$$

where the axis of real values is in the direction of the currents. The operating point for a particular value of δ can easily be fixed on the circle (Fig. 8) by the construction outlined in Section 9.3. From the impedance locus the current locus could be obtained by the method of inversion. This yields for the current locus a circle of the following parameters:

$$Centre: \quad \frac{V_1^2}{F^2 + G^2 - R^2}(F - jG)$$

$$Radius: \quad \frac{V_1 R}{F^2 + G^2 - R^2}$$

where the axis of real values is in the direction of the applied voltage.

It is obvious that the same expressions for the parameters of the current locus could have been obtained from eqn. (13), which gives the current in terms of s.

(5) THE APPLICATION OF THE EQUATIONS

The authors were unable to apply the theory to an actual machine and subject it to test. However, it is instructive to apply the equations to a motor whose constants have been given by Gibbs,[7] with the further assumption that $X_{33} = X_{12} = X_{23}$.

The values of ψ and ρ are so chosen that the quantity R_2 can reasonably be assumed to be constant from no load to full load. For a 120 h.p. 3-phase 50 c/s motor with a synchronous speed of 750 r.p.m. the constants are as follows:

$$R_1 = 0 \cdot 11 \text{ ohm} \qquad X_{1l} + X_{11} = 16 \cdot 45 \text{ ohms}$$
$$X_{21} + X_{22} = 16 \cdot 2 \text{ ohms}$$

$$X_{31} + X_{33} = 16 \cdot 2 \text{ ohms} \qquad X_{12} = X_{23} = 15 \cdot 7 \text{ ohms}$$
$$X_{13} = 16 \cdot 2 \text{ ohms}$$

$$r_{max} = 0 \cdot 52 \qquad R_2 + r^2R_3 = 0 \cdot 42 \text{ ohm}$$
$$V_1 = 400 \text{ volts}$$

For the assumption that $X_{33} = X_{23}$ the value of X_{3l} is 0·5 ohm.

(5.1) Performance at High Speed (for Small Values of ρ)

$\psi = 66°$

$\rho = -5°$

Steady losses (iron, friction, etc.) are assumed to be constant in the range considered and equal to $8 \cdot 2 \text{ kW}$.

Table 1 summarizes the calculations pertaining to input current, output, efficiency and power factor. The values as obtained by calculation by Gibbs[7] are shown in parenthesis for comparison. Figs. 3(a) and 3(b) show respectively the variation of gross output

Table 1

120 H.P. MOTOR: HIGH-SPEED CHARACTERISTICS

Percentage of full load output	100	72·4	47·2	24·4	4·08
Power factor	0·957 (lead) (1·0)	0·993 (lead) (0·978 lag)	0·953 (lag) (0·813 lag)	0·570 (lag) (0·485 lag)	0·184 (lag) (0·184 lag)
Percentage efficiency	83·4 (84·8)	83·7 (83·7)	80·8 (81·0)	70·1 (68·6)	24·9 (25·1)
Slip	−0·365 −0·4	−0·405 (−0·43)	−0·442 (−0·46)	−0·486 (−0·49)	−0·53 (−0·52)
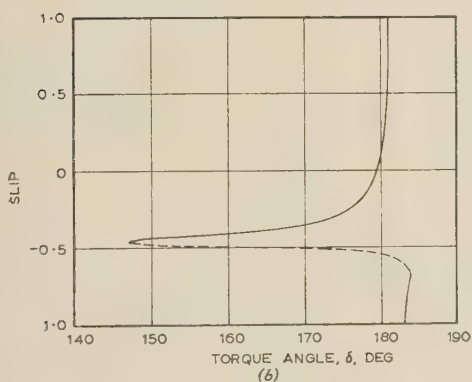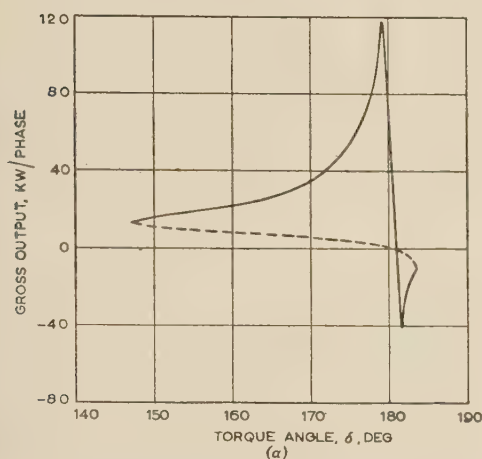
Fig. 3.—Performance curves for high-speed setting.

(a) Gross output/torque angle.
(b) Slip/torque angle.
———— Positive-root values.
---- Negative-root values.
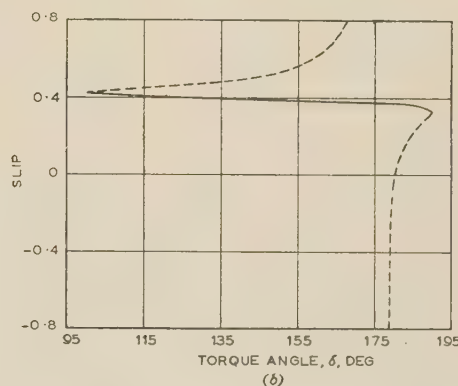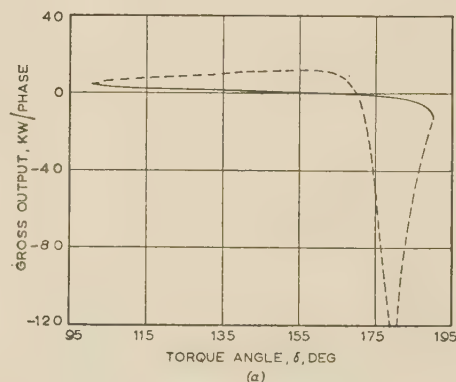
Fig. 5.—Performance curves for low-speed setting.

(a) Gross output/torque angle.
(b) Slip/torque angle.
———— Positive-root values.
---- Negative-root values.
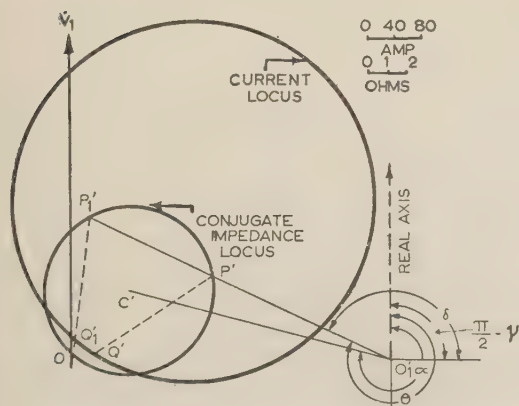
Fig. 4.—Current and conjugate impedance loci for high-speed setting.
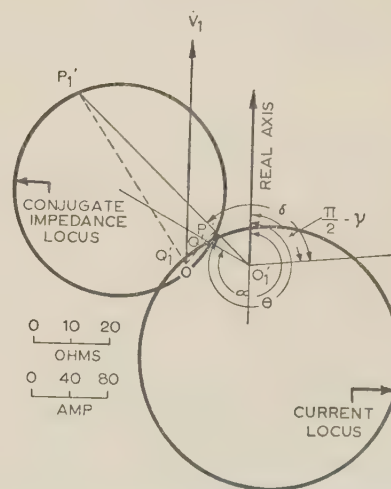
Fig. 6.—Current and conjugate impedance loci for low-speed setting.

Table 2

120 H.P. MOTOR: HIGH-SPEED CHARACTERISTICS

Percentage of full load output	96·4	87·5	67·5	37·3	11·9
Power factor	0·695 (lag) (0·62 lag)	0·734 (lag) (0·67 lag)	0·816 (lag) (0·754 lag)	0·948 (lag) (0·91 lag)	0·971 (lead) (0·89 lead)
Percentage efficiency	78·3 (76·2)	79·5 (77·8)	78·0 (76·4)	66·2 (66·6)	35·4 (45·5)
Slip	0·505 (0·533)	0·482 (0·50)	0·452 (0·46)	0·420 (0·42)	0·402 (0·39)

power and slip with the torque angle, δ. Fig. 4 gives the locus of the primary current and the point p corresponding to a torque angle of 155 electrical degrees.

(5.2) Performance at Low Speed (for Large Values of ρ)

$\psi = 55°$

$\rho = 172°$

Steady loss at low speed is $3 \cdot 4$ kW.

Table 2 and Figs. 5(a), 5(b) and (6) give the summary of the performance calculations.

(6) CONCLUSION

The equations of performance presented are more comprehensive than those offered by previous authors. In particular, the equations presented by Gibbs[7] can be readily deduced from the foregoing on the basis of the following simplifying assumptions:

(a) $X_{12} = X_{23}$.

(b) The reactance drop due to leakage fluxes in the tertiary winding is assumed to produce a voltage drop $I_2 X_{3l}$ in magnitude instead of $sI_2 X_{3l}$, which is actually the case.

The main difficulty in any predetermination is the difficulty of assigning definite values to the constants involved. This is particularly so for this machine as regards the value of $R_2 + r^2 R_3$, which is a function of the current due to the presence of the brushes. It is obvious, therefore, that the choice of the constants must be a matter of successive approximations based on the curve relating the resistance to the secondary current.

(7) ACKNOWLEDGMENT

The authors are grateful to the Director and the Department of Electrical Engineering of the Indian Institute of Technology, Kharagpur, for the facilities offered. In particular, they wish to thank Mr. H. E. Merchant for the services rendered.

(8) REFERENCES

(1) ARNOLD, A. H. M.: 'Circle-Diagram of Three-Phase Commutator Motor', *Journal I.E.E.*, 1926, **64**, p. 139.
(2) RUDRA, J. J.: 'Diagramme exacte et diagramme approximatif du moteur Schrage', *Revue Generale de L'Électricité*, 1934, **35**, p. 843.
(3) RUDRA, J. J.: 'Theorie generelise du moteur Schrage', *ibid.*, 1939, **46**, p. 49.
(4) CONRAD, A. G., ZWEIG, F., and CLARKE, J. G.: 'Theory of the Brush Shifting A.C. Motor', *Transactions of the American I.E.E.*, 1941, **60**, p. 829, and 1942, **61**, p. 502.
(5) KRON, G.: 'Application of Tensors to the Analysis of Rotating Electrical Machinery', *General Electric Review*, 1935, **38**, pp. 180–230, 282, 339, 386, 434, 473, 527, 580; 1936, **39**, pp. 108, 155, 201, 249, 297, 397 and 504; 1937, **40**, pp. 101, 197, 296, 389, 490 and 594.
(6) ADKINS, B., and GIBBS, W. J.: 'Polyphase Commutator Machines' (Cambridge University Press, 1951), p. 117.
(7) GIBBS, W. J.: 'The Equations and Circle Diagrams of the Schrage Motor', *Journal I.E.E.*, 1946, **93**, Part II, p. 621.
(8) KRON, G.: 'A Short Course in Tensor Analysis for Electrical Engineers' (Wiley, 1942).

(9) APPENDICES

(9.1) M.M.F. Relationships

(9.1.1) **For the Stator.**

Consider Fig. 1(a). The m.m.f. $I_{01}T_1$ required in winding 1 to produce the same flux linkage with itself as those due to all the m.m.f.'s F_1, F_2 and F_3 may be deduced from the following considerations:

The m.m.f. F_1 requires an ampere-turn in each phase of winding 1 which is given by $k_1 I_1 T_1 \varepsilon^{j\alpha}$, where $I_1 = k_1 I_1 \varepsilon^{j\alpha}$. The m.m.f. F_2 lags behind F_1 by δ and has a magnitude $k I_2 T_2$. This m.m.f. however, as seen from Fig. 2, acts on a magnetic circuit different from that corresponding to F_1. Consequently, the equivalent m.m.f. required in winding 1 to replace F_2 must have a magnitude $(X_{12}/X_{11})I_2 T_2$. Hence the ampere-turns required in winding 1 to produce the same flux linkages as F_2 is

$$\frac{X_{12}}{X_{11}} k_1 I_2 T_2 \varepsilon^{j(\alpha - \delta)}$$

By a similar argument it is seen that the equivalent m.m.f. to replace F_3 is

$$\frac{X_{13}}{X_{11}} r k_1 I_2 T_2 \varepsilon^{j(\alpha - \delta - \rho)}$$

Thus

$$I_{01}T_1 = k_1 I_1 T_1 \varepsilon^{j\alpha} + \left(\frac{X_{12}}{X_{11}}\right) k_1 I_2 T_2 \varepsilon^{j(\alpha - \delta)} + \left(\frac{X_{13}}{X_{11}}\right) k_1 r I_2 T_2 \varepsilon^{j(\alpha - \delta - \rho)} \quad \cdots \cdots (1$$

or $\quad I_{01} = I_1 \left[1 + \left(\frac{X_{12}}{X_{11}}\right)\frac{p}{k}\varepsilon^{-j\delta} + \left(\frac{X_{13}}{X_{11}}\right)\frac{p}{k}r\varepsilon^{-j(\delta + \rho)} \right]$

(9.1.2) **For the Rotor.**

In the case of the rotor it becomes necessary to consider, in the first instance, the m.m.f. relation for sub-synchronous and super-synchronous speeds separately.

(9.1.2.1) *Sub-synchronous operation* ($s > 0$).

As seen from the stator [Fig. 1(a)] F_1 leads F_2 by δ and F lags behind F_2 by ρ. The magnetic circuits acted upon by each of these m.m.f.'s are as shown in Fig. 2. Following the steps outlined in Section 9.1.1, it is seen that

$$I_{02}T_2 = k_2 \frac{X_{12}}{X_{22}} I_1 T_1 \varepsilon^{j(\beta + \delta)} + k_2 I_2 T_2 \varepsilon^{j\beta} + k_2 \frac{X_{32}}{X_{22}} r I_2 T_2 \varepsilon^{j(\beta - \rho)} \quad (2$$

where $\quad I_2 = k_2 I_2 \varepsilon^{j\beta}$

Hence $\quad \dot{I}_{02} = \dot{I}_2\left(1 + \dfrac{X_{12}}{X_{22}}\dfrac{k}{p}\varepsilon^{j\delta} + \dfrac{X_{32}}{X_{22}}r\dfrac{k}{p}\varepsilon^{-j\rho}\right)$

9.1.2.2) Super-synchronous operation ($s < 0$).

For super-synchronous operation, as seen from the stator [Fig. 1(a)], since the rotation of m.m.f.'s is now in the opposite direction, F_1 lags behind F_2 by δ and F_3 leads F_2 by ρ. Consequently

$\dot{I}_{02}T_2 = k_2\dfrac{X_{12}}{X_{22}}I_1T_1\varepsilon^{j(\beta-\delta)} + k_2I_2T_2\varepsilon^{j\beta} + k_2\dfrac{X_{32}}{X_{22}}rI_2T_2\varepsilon^{j(\beta+\rho)}$ (2b)

Hence $\quad \dot{I}_{02} = \dot{I}_2\left(1 + \dfrac{X_{12}}{X_{22}}\dfrac{k}{p}\varepsilon^{-j\delta} + \dfrac{X_{32}}{X_{22}}r\dfrac{k}{p}\varepsilon^{j\rho}\right)$

(9.2) E.M.F. Equation

For the assumption of negligible iron losses it is easily seen that the induced e.m.f.'s $\dot{V}_1'$ and $\dot{V}_2'$ in the respective windings 1 and 2 are given by the relation

$$\dot{V}_1' = j\dot{I}_{01}X_{11} = j\dot{I}_1\left[X_{11} + \varepsilon^{-j\delta}\dfrac{p}{k}(X_{12} + rX_{13}\varepsilon^{-j\rho})\right] \quad . \quad (3)$$

and $\quad \dot{V}_2' = j\dfrac{s}{k^2}\dot{I}_{02}X_{22} = j\dfrac{s\dot{I}_2}{k^2}\left(X_{12}\dfrac{k}{p}\varepsilon^{j\delta} + X_{22} + X_{32}r\varepsilon^{-j\rho}\right)$ (4a)

for sub-synchronous operation, or

$$= j\dfrac{s}{k^2}\dot{I}_2\left(X_{12}\dfrac{k}{p}\varepsilon^{-j\delta} + X_{22} + X_{32}r\varepsilon^{j\rho}\right) \quad . \quad . \quad (4b)$$

for super-synchronous operation.

It remains only to deduce the expression for the e.m.f. V_3' of winding 3 that is effective at the brushes.

Consider now the case of sub-synchronous operation for which the direction of rotation of the m.m.f.'s referred to the stator is as shown in Fig. 1(a). If the instantaneous current in winding 2 is $i_2 = \sqrt{2}I_2\cos(\beta + s\omega t)$, the axis of the m.m.f. F_2 leads that of this winding (in the direction of rotation of the m.m.f.'s) by an angle $(\beta + s\omega t)$ as shown in Fig. 7(a). The positions in space of the m.m.f.'s F_1, F_2 and F_3 are also indicated in the diagram. Since the rotor is rotating at an angular velocity $(1 - s)\omega$, the angular displacement of a fixed point P on the rotor at any instant t, with reference to the stationary axis of winding 2, may be written as

$$\alpha + (1 - s)\omega t$$

where α is the angular position of this point when t is zero. Consider a coil on the rotor with its axis at Q which is θ electrical degrees ahead of the fixed point P (in the direction of rotation of the m.m.f.'s). If Φ_{Q1}, Φ_{Q2} and Φ_{Q3} are the maximum fluxes linking this coil due to F_1, F_2, and F_3 respectively, the net flux Φ_Q is, as seen from Fig. 7(a), given by

$$\Phi_Q = \Phi_{Q1}\cos[\beta + s\omega t + \delta + \alpha + (1 - s)\omega t - \theta]$$
$$+ \Phi_{Q2}\cos[\beta + s\omega t + \alpha + (1 - s)\omega t - \theta]$$
$$+ \Phi_{Q3}\cos[\beta + s\omega t - \rho + \alpha + (1 - s)\omega t - \theta]$$

Consequently, the induced e.m.f. in this coil is

$$V_Q = -\dfrac{d}{dt}\Phi_Q = \omega[\Phi_{Q1}\sin(\beta + \delta + \alpha - \theta + \omega t)$$
$$+ \Phi_{Q2}\sin(\beta + \alpha - \theta + \omega t)$$
$$+ \Phi_{Q3}\sin(\beta + \alpha - \rho - \theta + \omega t)]$$
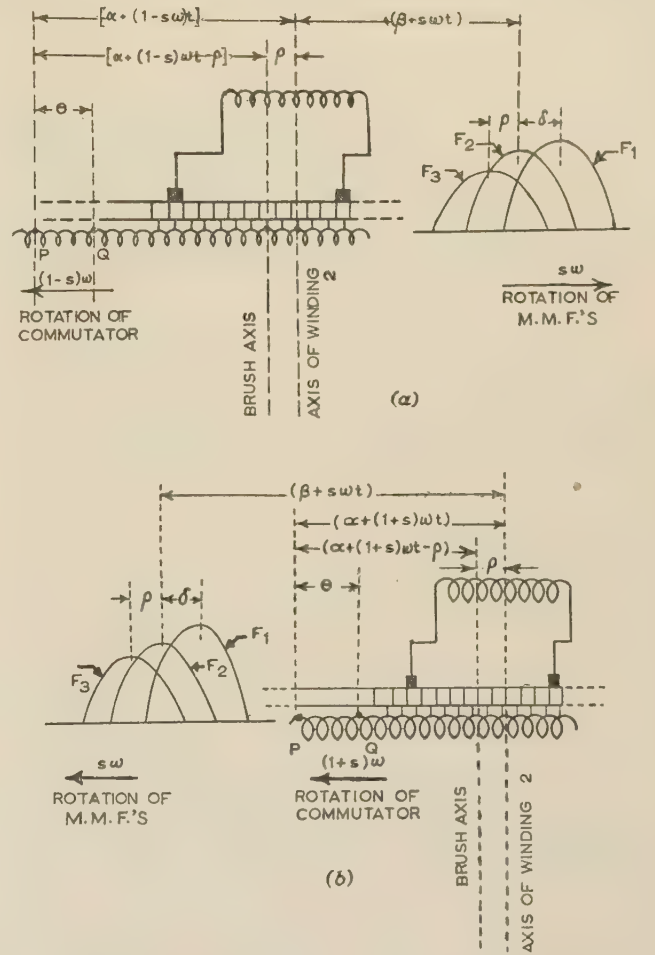
Fig. 7.—Instantaneous spatial positions of m.m.f.'s, of axis of winding 2 and of a specific coil in winding 3.
(a) For sub-synchronous operation.
(b) For super-synchronous operation.

If the winding between the brushes is replaced by an equivalent coil, then for this coil

$$\theta = [\alpha + (1 - s)\omega t - \rho]$$

and Φ_{Q1}, Φ_{Q2} and Φ_{Q3}, assume new magnitudes, say, Φ_{m1}, Φ_{m2} and Φ_{m3} respectively. Hence the e.m.f. V_3 collected between the brushes is given by

$$V_3 = (V_Q)_{\theta = \alpha + (1-s)\omega t - \rho}$$
$$= \omega[\Phi_{m1}\sin(\beta + \delta + \rho + s\omega t)$$
$$+ \Phi_{m2}\sin(\beta + s\omega t + \rho)$$
$$+ \Phi_{m3}\sin(\beta + s\omega t)]$$

From the definition for X_{33} it follows that

$$\sqrt{(2)}I_2\dfrac{r^2}{k^2}X_{33} = \omega\Phi_{m3}$$

Also $\quad \dfrac{\Phi_{m1}}{\Phi_{m3}} = \dfrac{I_1T_1}{rI_2T_2}\dfrac{X_{13}}{X_{33}}$ and $\dfrac{\Phi_{m2}}{\Phi_{m3}} = \left(\dfrac{I_2T_2}{rI_2T_2}\right)\dfrac{X_{23}}{X_{33}}$

Bearing in mind that the instantaneous current in winding 2 is $\sqrt{2}I_2\cos(\beta + s\omega t)$, the potential drop in winding 3 in the

direction of flow of I_2 is $-V_3$ and can be written in vector form as

$$\dot{V}_3' = j\frac{I_2}{k^2}\left(r\frac{k}{p}X_{13}\varepsilon^{j(\delta+\rho)} + rX_{23}\varepsilon^{j\rho} + r^2X_{33}\right) \quad . \quad (5a)$$

Consider next the super-synchronous operation for which the position of m.m.f.'s, P, Q, etc., are given in Fig. 7(b). Here also, if the instantaneous current in winding 2 is $i_2 = \sqrt{2}I_2 \cos(\beta + s\omega t)$, then, proceeding as before and bearing in mind that the angular velocity of the rotor is $(1 + s)\omega$, the net flux Φ_Q linked with a coil having its axis at Q is given by

$$\Phi_Q = \Phi_{Q1}\cos[\beta + s\omega t - \delta - \alpha - (1 + s)\omega t + \theta]$$
$$+ \Phi_{Q2}\cos[\beta + s\omega t - \alpha - (1 + s)\omega t + \theta]$$
$$+ \Phi_{Q3}\cos[\beta + s\omega t + \rho - \alpha - (1 + s)\omega t + \theta]$$

Proceeding as before, since θ for the equivalent coil replacing the winding between brushes is

$$\alpha + (1 + s)\omega t - \rho$$

the induced e.m.f. is

$$V_3 = -\omega[\Phi_{m1}\sin(\beta + s\omega t - \delta - \rho)$$
$$+ \Phi_{m2}\sin(\beta + s\omega t - \rho)$$
$$+ \Phi_{m3}\sin(\beta + s\omega t)]$$

whence

$$\dot{V}_3' = -j\frac{I_2}{k^2}\left[r\frac{k}{p}X_{13}\varepsilon^{-j(\delta+\rho)} + rX_{23}\varepsilon^{-j\rho} + r^2X_{33}\right] \quad (5b)$$

(9.3) Impedance and Current Circles

Consider the circle in Fig. 8, where C is the centre of the circle, O is the origin, O_1 a point in the plane of the paper, and
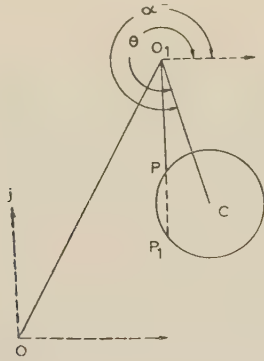
Fig. 8.—Impedance locus for the Schrage motor.

O_1PP_1 a secant making an angle α with the axis of real values in the counter-clockwise direction as shown. The line O_1C leads the axis of real values by an angle θ. Then

$$\overrightarrow{OP} = \overrightarrow{OO_1} + \overrightarrow{O_1P}$$
$$= \overrightarrow{OO_1} + O_1P\varepsilon^{j\theta}$$
$$= \overrightarrow{OO_1} + \left[O_1C(\cos\alpha\cos\theta + \sin\alpha\sin\theta)\right.$$
$$\left. \pm\{[O_1C(\cos\alpha\cos\theta + \sin\alpha\sin\theta)]^2 - [O_1C^2 - CP^2]\}\right]\varepsilon^{j\theta}$$

the two roots giving the points P and P_1 respectively. Th[e] impedance function is, from eqn. (12),

$$R_1 + j(X_{1l} + X_{11}) + j\varepsilon^{-j\delta}\left\{-\frac{h\sin\delta + \cos\delta}{2n}\right.$$
$$\left. \pm\sqrt{\left[\left(\frac{h\sin\delta + l\cos\delta}{2n}\right)^2 - \frac{m}{n}\right]}\right\}A\varepsilon^{-[}$$

where $A\varepsilon^{-j\nu} = X_{12} + rX_{13}\varepsilon^{-j\rho}$

i.e.

$$R_1 + j(X_{1l} + X_{11}) + \varepsilon^{j(\pi/2-\delta-\nu)}\left\{-\left(\frac{Ah\sin\delta + Al\cos\delta}{2n}\right)\right.$$
$$\left. \pm\sqrt{\left[\left(\frac{Ah\sin\delta + Al\cos\delta}{2n}\right)^2 - A^2\frac{m}{n}\right]}\right]$$

or

$$[R_1 + j(X_{1l} + X_{11})] + \varepsilon^{j\theta}[-B \pm \sqrt{(B^2 - A^2m/n)}]$$

where $B =$

$$+\frac{A(h\cos\nu + l\sin\nu)\cos\theta + A(-h\sin\nu + l\cos\nu)\sin\theta}{2n}$$

and

$$\theta = \frac{\pi}{2} - \nu - \delta$$

By comparison with Fig. 8, the impedance function defines [a] circle, where

$$\overrightarrow{OO_1} = R_1 + j(X_{1l} + X_{11})$$

$$O_1C^2 = \frac{A^2}{4n^2}(h^2 + l^2)$$

$$CP^2 = \frac{A^2}{4n^2}(h^2 + l^2 - 4mn)$$

and

$$\tan\alpha = \frac{-h\sin\nu + l\cos\nu}{h\cos\nu + l\sin\nu}$$

$$\overrightarrow{OC} = \overrightarrow{OO_1} + O_1C\cos\alpha + jO_1C\sin\alpha$$
$$= R_1 - \frac{A}{2n}(h\cos\nu + l\sin\nu)$$
$$+ j\left[X_{1l} + X_{11} - \frac{A}{2n}(-h\sin\nu + l\cos\nu)\right]$$

For positive values of

$$A^2\frac{m}{n} = O_1C^2 - CP^2$$

O_1 lies outside the circle. For this case, the operating points ar[e] possible for values of θ satisfying the inequality

$$\theta_1 < \theta < \theta_2$$

where $|O_1C\sin(\alpha - \theta_1)| = |O_1C\sin(\theta_2 - \alpha)| = CP$. There-fore the corresponding limits of θ are

$$\left(\frac{\pi}{2} - \nu - \theta_1\right) \text{ and } \left(\frac{\pi}{2} - \nu - \theta_2\right)$$

PROPERTIES OF SOME D.C.–A.C. CHOPPER CIRCUITS

By I. C. HUTCHEON, M.A., A.M.I.Mech.E., Associate Member.

SUMMARY

The paper analyses five basic types of chopper circuit suitable for converting low-level direct voltages to alternating voltages. Three of them are capacitance coupled and two are transformer coupled; all have resistive loads.

Both transformer-coupled circuits have make-before-break (overlap) contact action and incorporate a stopper resistor. If the overlap is small, this arrangement provides input resistances which approach those of the equivalent buffered underlap circuits, together with several other advantages.

Expressions are derived for gain, input resistance and optimum chopper resistance in terms of the component values.

LIST OF PRINCIPAL SYMBOLS

C_1 = Filter (reservoir) capacitance, farads.
C_2 = Coupling capacitance, farads.
C_T = Capacitance across transformer referred to primary, farads.
i = Instantaneous magnetizing current at time t, amp.
$i_a = i + i_b$, amp.
i_b = Instantaneous current in load, referred to primary, amp.
$-i_0$ = Value of i at $t = 0$, amp.
$+i_0$ = Value of i at $t = t_2$, amp.
L = Inductance of primary winding, henrys.
k = Secondary/primary turns ratio (circuit 4).
  = Secondary/half-primary turns ratio (circuit 5).
Q_2 = Net charge drawn from C_1 during t_2, coulombs.
R_1 = Filter resistance, ohms.
R_2 = Stopper resistance, ohms.
R_3 = Load resistance, ohms.
$R_3' = R_3/k^2$, ohms.
R_s = Secondary winding resistance, ohms.
$R_s' = R_s/k^2$, ohms.
R_p = Primary winding resistance (circuit 4), ohms.
R_c = Shunt resistance equivalent to C_T, ohms.
$R_a = R_2 + R_p$, ohms.
$R_b = R_3' + R_s'$, ohms.
R_0 = Input resistance less filter resistance, ohms.
$R_e = R_a R_b/(R_a + R_b)$, ohms.
t_1 = Chopping period, sec.
t_2 = Through or reverse period, sec.
t_3 = Overlap period (circuits 2 and 4), sec.
$t_4 = \frac{1}{2}t_1 - t_2 - 2t_3$ (circuit 4), sec.
  = Zero (circuit 5).
V = Direct input voltage, volts.
V_1 = Voltage across C_1, volts.
V_3 = Twice mean output voltage during t_2, volts.
$V_3'' = V_3/k$, volts.
v = Instantaneous output voltage, volts.
$v' = v/k$, volts.
ω = Angular chopping frequency = $2\pi/t_1$, rad/sec.

(1) INTRODUCTION

Mechanical choppers, or contact modulators, are currently in use on a fairly wide scale for applications which include:

(*a*) Conversion of low direct voltages to alternating voltages for easy drift-free amplification.
(*b*) Reconversion to d.c. of amplified a.c. signals.
(*c*) Stabilization of direct-coupled d.c. amplifiers.
(*d*) Isolation of d.c. circuits from one another by means of a transfer or 'bucket' capacitor.

One of the principal uses is (*a*), and commercial designs of chopper are available to provide voltage stability of 10^{-7} to 10^{-6} volt, and current stability of 10^{-12} to 10^{-11} amp. Even better performance has been obtained from choppers specially designed for research purposes. Other devices, such as the vibrating capacitor and the magnetic modulator, have their own fields of use which partly but not entirely overlap that of the mechanical chopper. The performance of d.c.–a.c. chopper circuits is consequently of considerable interest, and five circuits are analysed in the paper. They represent a much larger number of practical circuits differing only in minor details.

Three of the circuits are capacitance coupled, and their analysis is straightforward. The two other circuits are transformer coupled, with make-before-break ('overlap') contact settings and a stopper resistor which limits the current drain during the overlap periods. If the overlap is small, they can be made to give nearly as high an input resistance as the similar circuits with underlap contact settings and a buffer capacitor.[5] They also provide some features which are advantageous when very small signals are to be handled.

In all cases the degree of current and voltage stability finally obtained is largely dependent on the magnitude of the spurious signals introduced by the chopper. Stability also depends to some extent on the properties of the other components, on the circuit arrangement and on the magnitude of stray pick-up from external sources. These factors are discussed briefly, but chopper design is not dealt with in the paper.

(2) FUNCTIONAL ANALYSIS OF CIRCUITS

Most d.c.–a.c. chopper circuits, of which the five considered are typical, have a number of common features.

(2.1) Input Filter

The input voltage is generally applied through a low-pass RC or LC filter, the capacitor of which acts as a reservoir to supply the pulsating currents drawn by the switching section, and ensures that a steady current is taken from the source: this minimizes the effect of source resistance. The filter also attenuates spurious alternating voltages which arrive with the d.c. signal. Single- or multi-stage filters may be used, and perform identically under equilibrium conditions, only the total d.c. resistance being significant. LC filters are more efficient, but the inductors require heavy shielding against stray magnetic fields, and are therefore seldom used.

(2.2) Switching Circuit

The filter capacitor is followed by a switching circuit which may be

(a) On-off.
(b) Single-pole double-throw break-before-make (underlap).
(c) Single-pole double-throw make-before-break (overlap).
(d) Double-pole change-over (underlap or overlap).

Capacitance-coupled circuits may make use of any of the first three arrangements, and the typical circuits shown in Figs. 1(a), (b) and (c) are based on them. The fourth arrangement is
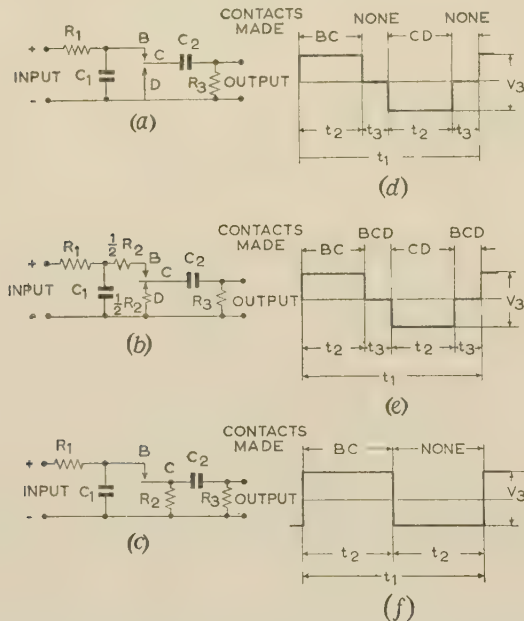
Fig. 1.—Capacitance-coupled chopper circuits and waveforms.

    (a) Single-pole double-throw with underlap (circuit 1).
    (b) Single-pole double-throw with overlap (circuit 2).
    (c) On–off (circuit 3).
    (d) Output voltage waveform for circuit 1.
    (e) Output voltage waveform for circuit 2.
    (f) Output voltage waveform for circuit 3.

generally not suitable for capacitance coupling, because of earthing problems.

Any of the arrangements can be used with transformer coupling, but we are here concerned only with the overlap cases, i.e. (c) and (d).

Other circuits, such as a bucket-capacitor circuit for isolation purposes, may be combined with any of the above, but do not alter the principle of their operation.

(2.3) Stopper Resistor

When there is a make-before-break contact action the filter capacitor is short-circuited during each overlap period. Consequently a stopper resistor must be included between the filter capacitor and the chopper to limit the discharge from the former.

If the stopper resistance is varied, the gain passes through a maximum, the optimum resistance depending on the other circuit values. For reasons of symmetry, the stopper is often split into two equal parts.

(2.4) Output Coupling

The switch is coupled to the load by either capacitors or a transformer. Capacitance-coupled circuits can be made to have high input resistances, and transformer coupling is used to increase voltage sensitivity at the expense of input resistance, or to provide isolation of input from output. Complete d.c.

isolation is obtained by using a transformer or two coupling capacitors.

Transformer coupling provides a degree of a.c. isolation dependent upon the inter-winding capacitance [Fig. 5(a)], which is made as small as possible. A.C. isolation in capacitance-coupled circuits is less practicable unless bucket-capacitor circuits are used.[2, 14]

(2.5) Load

The final load is normally resistive, e.g. the grid resistor of a valve. In transformer-coupled circuits the core-loss resistance is in parallel with the load, and must be taken into account when the load is high; it is sufficiently accurate to take the loss resistance as measured with sine-wave inputs.

The effect of shunt capacitance, e.g. winding self-capacitance will be considered separately.

(2.6) More than One Input

The voltage to be chopped is often the difference between two or more larger voltages, which can be fed into any convenient parts of the d.c. circuit before the switch. Usually the measured voltage is applied to the filter, while the balancing voltage is fed in after the filter capacitor, as shown in Fig. 5.

In determining circuit performance it is convenient to assume that there is a single input only, equal to the difference between the two or more actual inputs. This causes no loss of generality provided that the voltages are steady.

(3) SOME BASIC REQUIREMENTS AND DEFINITIONS

Some of the less desirable circuits can be eliminated immediately by consideration of requirements for symmetry.

The remaining circuits can then be evaluated in the light of their properties as found by analysis. Two important properties, the circuit gain and input resistance, are needed for this, and are defined below.

(3.1) Symmetry

Whenever possible, chopping circuits should provide symmetrical switching, i.e. the through and reverse circuit configurations should be similar and of equal duration. This ensures that any pick-up at the chopping frequency is converted so far as possible into higher harmonics to which demodulating circuits can be made unresponsive. It also assists the cancelling of thermal e.m.f.'s generated in the circuit. On-off switching is bad in this respect.

The output impedance should be the same during alternate half-cycles, otherwise hum in the following valve may be modified undesirably (e.g. 100 c/s hum might be modified to 50 c/s, which would cause errors with 50 c/s chopping), and in capacitance-coupled circuits, slow variations in grid current of the following valve (due, for example, to supply-voltage variations) may be chopped and so produce an a.c. output. (A high-pass filter is often included between the chopping circuit and the valve to reduce this effect.)[9]

The output impedance also should not rise to high values during switching, otherwise the circuit becomes susceptible to pick-up effects and transients may occur in the output signal. Underlap contact settings are bad in this respect.

(3.2) Gain

When the waveform is rectangular the gain is defined as the ratio between the peak-to-peak a.c. output signal and the d.c. input. If the output waveform has appreciable droop, as is the case with transformer-coupled circuits, it is convenient to define the a.c. output as being twice the mean height of the waveform

during the through or reverse period t_2. This definition has been found quite satisfactory in practice.

(3.3) Input Resistance

All the circuits draw a pulsating current from the input capacitor, which is assumed to be large, and an equivalent steady current flows through the filter resistor. The input voltage

Typical output waveforms are also shown in Figs. 1(d), (e) and (f), which also give the terminology for the switching times.

The method of analysis is similar for all three circuits, and is typified by the treatment of circuit 2 given in the following Section. Expressions for gain, input resistance, optimum stopper resistance, etc., are given in Table 1, and these have been checked experimentally.

Table 1

PARAMETERS OF CAPACITANCE-COUPLED CHOPPER CIRCUITS WITH C_1 AND C_2 INFINITE

Parameter	Circuit 1	Circuit 2	Circuit 3
Gain	$1 - \dfrac{1}{1 + 2\dfrac{t_1}{t_2}\dfrac{R_3}{R_1}}$	$\dfrac{1}{1 + 2\dfrac{t_3}{t_1}\dfrac{R_1}{R_2} + \dfrac{1}{2}\dfrac{R_2}{R_3} + \dfrac{1}{2}\left(\dfrac{t_1 - t_2}{t_1}\right)\dfrac{R_1}{R_3}}$	$\dfrac{1}{1 + \dfrac{1}{2}\dfrac{R_1}{R_2} + \dfrac{1}{2}\dfrac{R_2}{R_3} + \dfrac{1}{2}\dfrac{R_1}{R_3}}$
Input resistance	$R_1 + 2\dfrac{t_1}{t_2}R_3$	$R_1 + R_2\left[\dfrac{2 + \dfrac{R_2}{R_3}}{\left(\dfrac{t_1 - t_2}{t_1}\right)\dfrac{R_2}{R_3} + 4\dfrac{t_3}{t_1}}\right]$	$R_1 + R_2 + \dfrac{R_2 R_3}{R_2 + R_3}$
$R_{2(opt)}$ for maximum gain		$2\sqrt{\left(\dfrac{t_3}{t_1}R_1 R_3\right)}$	$\sqrt{(R_1 R_3)}$
Maximum gain		$\dfrac{1}{1 + 2\sqrt{\left(\dfrac{t_3}{t_1}\dfrac{R_1}{R_3}\right)} + \dfrac{1}{2}\left(\dfrac{t_1 - t_2}{t_1}\right)\dfrac{R_1}{R_3}}$	$\dfrac{1}{1 + \sqrt{\dfrac{R_1}{R_3}} + \dfrac{1}{2}\dfrac{R_1}{R_3}}$
Input resistance with $R_{2(opt)}$		$R_1 + R_1\left[\dfrac{1 + \sqrt{\left(\dfrac{t_3}{t_1}\dfrac{R_1}{R_3}\right)}}{\dfrac{1}{2}\left(\dfrac{t_1 - t_2}{t_1}\right)\dfrac{R_1}{R_3} + \sqrt{\left(\dfrac{t_3}{t_1}\dfrac{R_1}{R_3}\right)}}\right]$	$R_1 + R_1\sqrt{\dfrac{R_3}{R_1}}\left[1 + \dfrac{1}{1 + \sqrt{\dfrac{R_1}{R_3}}}\right]$

divided by this current gives the input resistance of the circuit, which is made up of the filter resistance, R_1, in series with a resistance R_0 denoting the rest of the circuit.

If R_1 is considered as variable, the gain A is given by $A = KR_0/(R_1 + R_0)$, where K is a constant, and is halved if R_1 is increased by an additional amount $R_1 + R_0$. The input resistance can therefore be defined as that value of additional series resistance which halves the output signal.

Differentiation of the above expression gives the following more general relationships which enable the input resistance to be derived from the gain equation:

$$\text{Input resistance} = \frac{-A}{\partial A/\partial R_1} = \frac{1/A}{\partial(1/A)/\partial R_1}$$

It should be noted that, since R_0 is treated as constant, the stopper resistance (if any), on which R_0 depends, must appear explicitly in the gain equation and be treated as constant.

(4) CAPACITANCE-COUPLED CIRCUITS

The three basic capacitance-coupled circuits are shown in Figs. 1(a), (b) and (c): the first has equal through and reverse periods with variable contact underlap, and the second is similar, but has overlap and a split stopper resistance; both can be made fully symmetrical by replacing C_1 with two capacitors in series and taking the lower output connection from their junction, at the same time replacing R_1 with a resistor in each input lead. The third (on-off) is asymmetrical and less efficient, but is included because it represents a circuit which has been considerably used in practice[14] (this has a bucket capacitor which provides both isolating and on-off chopping actions).

(4.1) Analysis of Circuit 2 [Fig. 1(b)]

It is assumed that C_1 and C_2 are large and therefore attain steady potentials, V_1 and V_2, where V_1 is rather less than the input voltage, V. Consideration of the charges flowing into and out of C_2, which must add up to zero over one cycle, shows that in this case $V_1 = 2V_2$.

The peak-to-peak output of the circuit is

$$V_3 = V_1 \frac{R_3}{R_3 + \frac{1}{2}R_2} \quad\quad . \quad . \quad . \quad . \quad (1)$$

The charges flowing into and out of C_1 must also add up to zero over a cycle, giving

$$t_1\left(\frac{V - V_1}{R_1}\right) = \frac{t_2(V_1 - V_2)}{R_3 + \frac{1}{2}R_2} + 2t_3\frac{V_1}{R_2} \quad . \quad . \quad (2)$$

(No current flows through the connection to C_2 during the overlap period.) Solving eqn. (2) for V_1 in terms of V, and inserting the result in eqn. (1) gives the following expressions for circuit gain:

$$\frac{1}{\text{Gain}} = \frac{V}{V_3} = 1 + \frac{2t_3}{t_1}\frac{R_1}{R_2} + \frac{1}{2}\frac{R_2}{R_3} + \frac{(t_1 - t_2)}{2t_1}\frac{R_1}{R_3} \quad . \quad (3)$$

The value of R_2 giving greatest gain is found by differentiating with respect to R_2, giving

$$R_{2(opt)} = 2\sqrt{\left(\frac{t_3}{t_1}R_1 R_3\right)} \quad . \quad . \quad . \quad . \quad (4)$$

The input resistance is obtained from eqn. (3) in which R_2 appears and is therefore treated as constant; thus

$$\text{Input resistance} = \frac{1}{\text{Gain}} \div \frac{\delta(1/\text{gain})}{\delta R_1}$$

$$= R_1 + R_2 \left[\frac{2 \div \dfrac{R_2}{R_3}}{\left(\dfrac{t_1 - t_2}{t_1}\right)\dfrac{R_2}{R_3} + 4\dfrac{t_3}{t_1}} \right] \quad . \quad . \quad . \quad (5)$$

The optimum value of R_2 given by eqn. (4) can be inserted into eqns. (3) and (5) to give the maximum gain and the corresponding input resistance. (The latter does not have a maximum, but increases continually with R_2.)

(4.2) Approximations

If C_1 is infinite, the expressions for the three circuits are exact. Often they can be simplified to suit particular problems in which particular assumptions can be made. For example, if the chopper design is such that t_3 is small, the input resistance of circuit 1 is approximately $R_1 + 4R_3$.

In circuit 2, if t_3 is small, so is R_2; the input resistance is then given by the sum of R_1 and the parallel resultant of $4R_3$ and $R_2(t_1/2t_3)$. This can also be seen directly by inspection of the circuit. If the value of the parallel resultant is R_0, the gain is approximately $R_0/(R_0 + R_1)$.

Approximate expressions are not listed, since they are best derived as required.

(4.3) Discussion

If the overlap time, t_3, can be made small, circuit 2 approaches circuit 1 in efficiency. It is then often to be preferred, because the output impedance does not rise to high values during switching. The output impedance during t_2 is largely determined by the stopper resistance, and hence by t_3, which should be small.

Circuit 3 is less efficient than the other two and would be used only if the additional contact were not available, or perhaps to give greater reliability.

Many of the expressions can be plotted against R_1/R_3 for given switching times, thus enabling suitable values of this ratio to be chosen rapidly for circuits using a particular chopper.

(5) TRANSFORMER-COUPLED CIRCUITS

Two transformer-coupled circuits are considered, both being symmetrical and having overlap contact actions which ensure that the transformer primary is never open-circuited. They are a double-pole change-over single-primary circuit (No. 4) and a single-pole double-throw double-primary circuit (No. 5). They operate in a similar way, and the first will be considered in detail; the results will then be applied to the second.

(5.1) Operation of Double-Pole Change-Over Circuit (No. 4)

The circuit, equivalent circuit and typical waveforms are shown in Figs. 2(a), 2(c) and 3 respectively, and the contact arrangement of one suitable chopper is shown in Fig. 4.

When transient phenomena have disappeared during the first few cycles after switching on, the large filter capacitor C_1 has a steady potential V_1 which is rather less than the input, V.

At time $t = 0$ the capacitor is connected via R_2 to the transformer, and currents i_a, i and i_b flow, varying with time. After a time t_2 the current in L has reached a value $+i_0$ and the transformer is short-circuited for a time $t_4 + 2t_3$. During this period
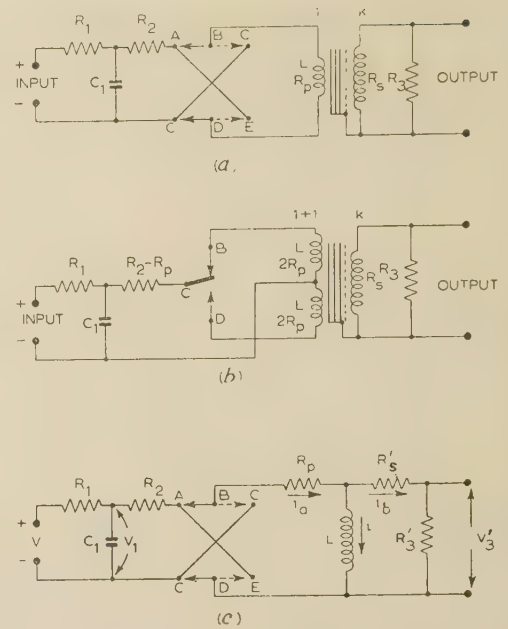
Fig. 2.—Transformer-coupled chopper circuits.

(a) Double-pole change-over with overlap (circuit 4).
(b) Single-pole double-throw with overlap (circuit 5).
(c) Equivalent for (a) and (b).
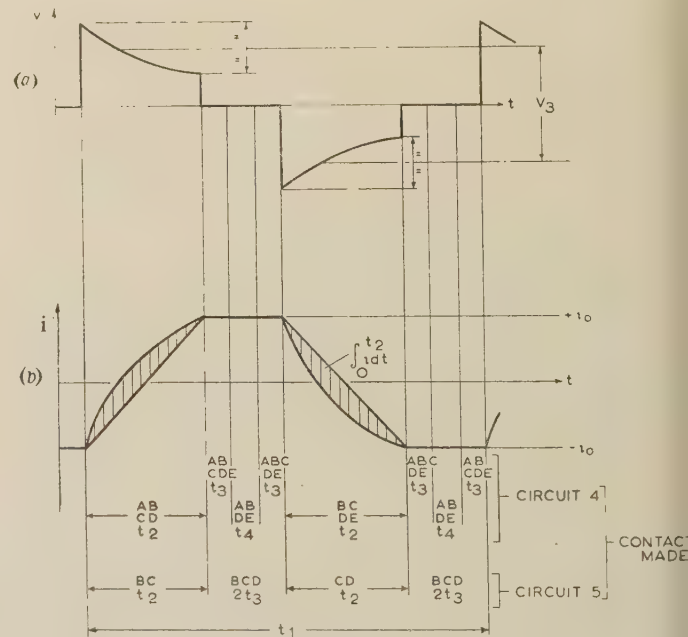
Fig. 3.—Waveforms for transformer-coupled circuits.

(a) Output voltage.
(b) Magnetizing current.
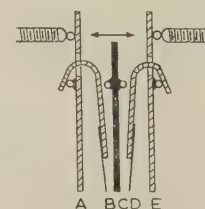
Fig. 4.—Contact arrangement.

the flux in the transformer core is maintained by i_0, which does not die away appreciably since the primary resistance is low. Consequently, when the next half-cycle commences and the capacitor is connected via R_2 to the transformer reversed, the initial current in L is $+i_0$; it then falls, during the period t_2, to $-i_0$.

With this information about the limits of current, an analysis of one half-cycle enables the currents and voltages at any time t to be found in terms of V_1 and the circuit characteristics. We can then find the net charge drawn per complete half-cycle from C_1, which is

$$\int_0^{t_2} (i + i_b)dt + 2t_3 V_1/R_2 \quad . \quad . \quad . \quad . \quad . \quad (6)$$

the second term being the charge drained from C_1 during the overlap periods t_3, when C_1 is short-circuited through R_2. Equating this to the charge flowing into C_1 per half-cycle gives V_1. Hence the gain can be found.

We can then find the input resistance, the value of R_2 which makes the gain a maximum, the corresponding value of the input resistance, and so on.

(5.2) Analysis of Circuit 4: General Case

The equivalent circuit is shown in Fig. 2(c).

Let
$$R_2 + R_p = R_a$$
$$R'_s + R'_3 = R_b$$
$$R_e = \frac{R_a R_b}{R_a + R_b}$$

Then the basic circuit equations for the period t_2 are

$$V_1 = R_a i_a + L \, di/dt \quad . \quad . \quad . \quad . \quad (7)$$

$$L \, di/dt = R_b i_b \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$i_a = i + i_b \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Eliminating i_a and i_b, integrating, and remembering that $i = -i_0$ when $t = 0$ and $i = +i_0$ when $t = t_2$, gives

$$i = \frac{V_1}{R_a}\left(1 - \frac{2\varepsilon^{-R_e t/L}}{1 + \varepsilon^{-R_e t_2/L}}\right) \quad . \quad . \quad . \quad (10)$$

The charge drawn from C_1 during t_2 is

$$Q_2 = \int_0^{t_2}(i + i_b)dt = \int_0^{t_2} i \, dt + \frac{L}{R_b}[i]_0^{t_2}$$

$$= \frac{V_1}{R_a}\left[t_2 - 2\frac{L}{R_a}\left(\frac{1 - \varepsilon^{-R_e t_2/L}}{1 + \varepsilon^{-R_e t_2/L}}\right)\right]$$

$$= \frac{V_1}{R_a}\left(t_2 - 2\frac{L}{R_a}\tanh\frac{1}{2}\frac{R_e t_2}{L}\right) \quad . \quad . \quad (11)$$

For convenience the tanh term is now replaced by the first two terms of its expansion as a power series. (The consequent error in Q_2 is less than 1% if $R_e t_2/L \not> 1$, and is about 12·5% if $R_e t_2/L = 2$. Values greater than unity are of little interest because they imply a peaky output waveform with large droop. With $R_e t_2/L = 1$ the droop is already 63%.)

Thus $$Q_2 = \frac{V_1}{R_a}\left\{t_2 - 2\frac{L}{R_a}\left[\frac{1}{2}\frac{R_e t_2}{L} - \frac{1}{3}\left(\frac{1}{2}\frac{R_e t_2}{L}\right)^3\right]\right\}$$

$$= \frac{V_1}{R_a}\left(\frac{R_a t_2}{R_a + R_b} + \frac{1}{12}\frac{R_e^3 t_2^3}{R_a L^2}\right) \quad . \quad . \quad . \quad (12)$$

We are now in a position to equate $Q_2 + 2t_3 V_1/R_2$, the charge drawn from C_1 during a complete half-cycle, to the charge flowing in during the same time, namely $t_1(V - V_1)/2R_1$. This gives the following expression for V_1:

$$\frac{V}{V_1} = 1 + 2\frac{R_1}{t_1}\left(2\frac{t_3}{R_2} + \frac{t_2}{R_a + R_b} + \frac{1}{12}\frac{R_e^3}{R_a^2}\frac{t_2^3}{L^2}\right) \quad . \quad (13)$$

We can now find the gain. The voltage across L at any time during t_2 is

$$L\frac{di}{dt} = 2V_1\frac{R_e}{R_a}\left(\frac{\varepsilon^{-R_e t/L}}{1 + \varepsilon^{-R_e t_2/L}}\right) \quad . \quad . \quad . \quad (14)$$

And the sum of the voltages at $t = 0$ and $t = t_2$ is

$$2V_1 R_e/R_a \quad . \quad . \quad . \quad . \quad . \quad (15)$$

This must be multiplied by R'_3/R_b to give the output across R'_3. The gain, as previously defined, is therefore

$$2k\frac{V_1}{V}\frac{R'_3}{(R_a + R_b)} \quad . \quad . \quad . \quad . \quad (16)$$

Inserting the value of V_1 given in eqn. (13) gives

$$\text{Gain} = \cfrac{2k}{\cfrac{R_a + R_b}{R'_3} + \cfrac{2}{t_1}\cfrac{R_1}{R'_3}\left[2t_3\left(\cfrac{R_a + R_b}{R_2}\right) + t_2 + \cfrac{1}{12}\cfrac{R_e^2 R_b}{R_a}\cfrac{t_2^3}{L^2}\right]} \quad . \quad . \quad . \quad (17)$$

The input resistance, $\dfrac{1}{\text{Gain}} \div \dfrac{\delta(1/\text{gain})}{\delta R_1}$ is thus

$$R_1 + \cfrac{R_a + R_b}{\cfrac{2}{t_1}\left[2t_3\left(\cfrac{R_a + R_b}{R_2}\right) + t_2 + \cfrac{1}{12}\cfrac{R_e^2 R_b}{R_a}\cfrac{t_2^3}{L^2}\right]} \quad . \quad (18)$$

And the gain is a maximum, with R_2 varying, when

$$2t_3\frac{(R_b + R_p)}{R_2^2} = \frac{1}{2}\frac{t_1}{R_1} + \frac{1}{12}\frac{t_2^3}{L^2}(R_b - R_a)\left(\frac{R_b}{R_a + R_b}\right)^3 . \quad (19)$$

(5.3) Approximations for Low-Resistance Case

The expressions developed so far are accurate to 1% for all component values, with the simple limitations that $R_e t_2/L \le 1$, and that C_1 can be considered to be infinite. However, they are too complicated for most purposes. Therefore we shall derive some simpler expressions which are approximately true when certain simplifying assumptions are made.

The first and most obvious step is to assume that L is infinite when the circuit resistances are relatively low. Inspection of the equations shows that the simplified expressions for gain and input resistance will be accurate to $100/K\%$ or better if

$$\left(\frac{L}{t_2}\right)^2 = \frac{1}{12}\frac{KR_e^3}{R_a\left(\frac{R_e}{R_b} + 2\frac{t_3}{t_2}\frac{R_a}{R_2}\right)} \quad . \quad . \quad . \quad (20)$$

(In practice it is necessary to assume that the low-resistance equations apply, use them to determine R_2 and then check that the above condition has been met.)

Putting $L = \infty$ we find that

$$Q_2 = V_1 t_2/(R_a + R_b) \quad . \quad . \quad . \quad (21)$$

which is obviously correct. Other simplified expressions are listed in Table 2. The expression for $R_{2(opt)}$ is sufficiently simple to be inserted into the expressions for gain and input resistance, giving the maximum gain and the corresponding input resistance

Table 2

PARAMETERS FOR CIRCUITS 4 AND 5 WITH C_1 INFINITE

Parameter	Low-resistance case	High-resistance case
Assumptions	$$\left(\frac{L}{t_2}\right)^2 \geqslant K \frac{R_e^3}{12R_a\left(\frac{R_e}{R_b} + 2\frac{t_3}{t_2}\frac{R_a}{R_2}\right)}$$ (Consequent errors in gain and input resistance are below $100/K$ %)	$R_2 \leqslant L/t_2$ (consequent errors in gain and input resistance are less than 1%) $\left.\begin{array}{l} R_3' \geqslant KR_s' \\ R_3' \geqslant KR_2 \\ R_2 \geqslant KR_p \end{array}\right\}$ (Consequent errors in gain and input resistance are about $100/K$%)
Gain	$$\frac{2k}{\frac{R_a}{R_3'} + \frac{R_b}{R_3'} + 2\frac{R_1}{R_3'}\left[\frac{t_2}{t_1} + 2\frac{t_3}{t_1}\left(\frac{R_a + R_b}{R_2}\right)\right]}$$	$$\frac{2k}{1 + \frac{R_1}{t_1}\left(4\frac{t_3}{R_2} + 2\frac{t_2}{R_3'} + \frac{1}{6}\frac{t_2^3 R_2}{L^2}\right)}$$
Input resistance	$$R_1 + \frac{1}{2}\left[\frac{R_a + R_b}{\frac{t_2}{t_1} + 2\frac{t_3}{t_1}\left(\frac{R_a + R_b}{R_2}\right)}\right]$$	$$R_1 + \left(\frac{R_2}{4\frac{t_3}{t_1} + 2\frac{t_2}{t_1}\frac{R_2}{R_3'} + \frac{1}{6}\frac{t_2^3}{t_1}\frac{R_2^2}{L^2}}\right)$$
$R_{2(opt)}$ for maximum gain	$$2\sqrt{\left[\frac{t_3}{t_1}R_1(R_b + R_p)\right]}$$	$$\sqrt{\frac{24t_3}{\left(\frac{t_2^3}{L^2} + 6\frac{t_1}{R_1 R_3'}\right)}}^\dagger \simeq \frac{L}{t_2}\sqrt{\frac{24t_3}{t_2}}^{*\dagger}$$
Maximum gain	$$\frac{2k}{\frac{R_b + R_p}{R_3'} + 2\left(\frac{t_2 + 2t_3}{t_1}\right)\frac{R_1}{R_3'} + \frac{4}{R_3'}\sqrt{\left[\frac{t_3}{t_1}R_1(R_b + R_p)\right]}}$$	$$\frac{2k}{1 + 2\frac{R_1}{t_1}\left[\sqrt{\left(\frac{2}{3}\right)\frac{t_2}{L}\sqrt{(t_2 t_3)}} + \frac{t_2}{R_3'}\right]}^*$$
Input resistance with $R_{2(opt)}$	$$R_1 + \frac{1}{2}\left\{\frac{R_b + R_p + 2\sqrt{\left[\frac{t_3}{t_1}(R_b + R_p)R_1\right]}}{\frac{t_2 + 2t_3}{t_1} + \sqrt{\left[\frac{t_3}{t_1}\frac{(R_b + R_p)}{R_1}\right]}}\right\}$$	$$R_1 + \frac{1}{2}\left[\frac{t_1/t_2}{\sqrt{\left(\frac{2}{3}\right)\frac{1}{L}\sqrt{(t_2 t_3)}} + \frac{1}{R_3'}}\right]^{*\dagger}$$
Term describing output waveform	$\varepsilon^{-R_e t/L}$	$\varepsilon^{-R_2 t/L}$

* $R_1 \geqslant 6\frac{t_1}{t_2}\left(\frac{L}{t_2}\right)^2\frac{1}{R_3'}$ assumed additionally.

† In these cases the input resistance is also a maximum.

(which is not a maximum). Most of the expressions can be plotted against R_1/R_3 for particular values of winding resistance and chopper characteristics.

(5.4) Approximations for High-Resistance Case

The second set of simple approximate equations is developed on the assumption that R_3 is very large. In practice this usually means that the transformer is terminated by a valve with no grid leak. R_3 is then the valve input resistance in parallel with the core-loss resistance of the transformer. We thus assume

$$R_b (= R_3' + R_s') = R_3'$$
$$R_a (= R_2 + R_p) = R_2$$
$$R_b + R_a = R_b$$

which implies that $R_3' \geqslant KR_s'$, $R_3' \geqslant KR_2$ and $R_2 \geqslant KR_p$, where K is suitably large, e.g. 10 or more depending on the accuracy required.

The first implication is very well met, the terminating resistance being far greater than that of the transformer secondary winding. The second and third implications depend on R_2. It will be seen that the optimum value of R_2 under these conditions is, in fact, usually such as to exceed the primary resistance by more than 10 times, while being more than 10 times less than the terminating resistance. Some typical values are given in Section 6.1.

The gain equation (17) then simplifies to

$$\text{Gain} = \frac{2k}{1 + \frac{R_1}{t_1}\left(4\frac{t_3}{R_2} + 2\frac{t_2}{R_3'} + \frac{1}{6}\frac{t_2^3 R_2}{L^2}\right)} \quad . \quad . \quad (22)$$

the input-resistance equation (18) becomes

$$\text{Input resistance} = R_1 + \left(\frac{R_2}{4\frac{t_3}{t_1} + 2\frac{t_2}{t_1}\frac{R_2}{R_3'} + \frac{1}{6}\frac{t_2^3}{t_1}\frac{R_2^2}{L^2}}\right) \quad . \quad (23)$$

and the equation for $R_{2(opt)}$, (19), simplifies to

$$R_{2(opt)} = 2\sqrt{\frac{t_3}{\frac{1}{6}\frac{t_2^3}{L^2} + \frac{t_1}{R_1 R_3'}}} \quad . \quad . \quad . \quad (24)$$

It is interesting to note that the charge drawn during t_2 becomes

$$Q_2 = \frac{1}{12}\frac{V_1 R_2 t_2^3}{L^2} \quad . \quad . \quad . \quad . \quad (25)$$

which increases with R_2. This is analogous to the energy loss in an LC oscillator circuit due to resistance in series with the inductance.

(a)

(b)
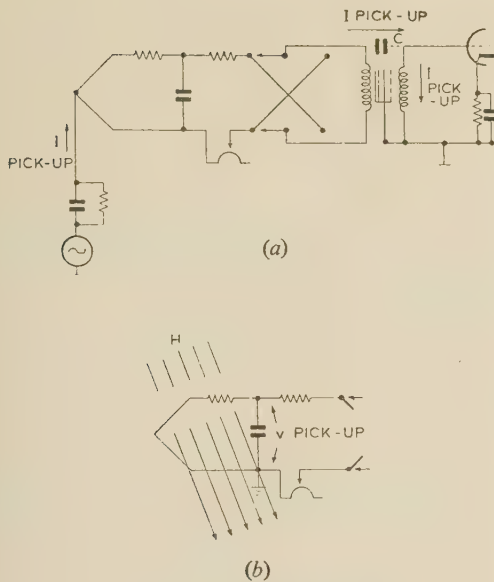
Fig. 5.—Types of pick-up.

(a) Leakage and capacitance.
(b) Magnetic.

Thus $R_{2(opt)}$ is low when R_1 is low, but rises to a limit dependent on L as R_1 is increased. In practice, R_1 comprises the filter resistance plus the resistance of the external source and their combined value is often so high as to require the upper-limit value for R_2. To render the circuit flexible for use with low and high values of R_1 the limiting value of R_2 would normally be used. We therefore make the further assumption that

$$R_1 R_3' \geqslant 6 \frac{t_1}{t_2} \frac{L^2}{t_2^2} \quad . \quad . \quad . \quad . \quad (26)$$

which enables us to simplify eqn. (24) to

$$R_{2(opt)} = \frac{L}{t_2} \sqrt{\frac{24 t_3}{t_2}} \quad . \quad . \quad . \quad . \quad (27)$$

This expression is independent of all resistance values. It can be inserted in eqns. (22) and (23) to give expressions for maximum gain and the corresponding input resistance which, in this case, is simultaneously a maximum (see Fig. 6).

It is interesting to note, too, that this value of R_2 makes Q_2 equal to the charge drawn by the short-circuit action during switching, i.e. to $2 t_3 V_1/R_2$.

(5.5) Single-Pole Double-Throw Circuit (No. 5)

This circuit [Fig. 2(b)] behaves in a manner similar to the double-pole change-over circuit. During t_2 the current in half the primary rises from a negative to an equal positive value ($\pm i_0$). Then all three contacts are connected, and no potential can appear across the two ends of the primary. Hence the core flux remains almost constant, being maintained by a current $i_0/2$ which circulates through the whole primary. When the first half of the primary is disconnected the current in the second half immediately rises to i_0, falling during t_2 to $-i_0$ again. While all the contacts are connected an additional current flows from C_1, through the two halves of the primary in opposite directions. This sets up no flux and no back-e.m.f. Hence C_1 is effectively short-circuited through the stopper resistance and the two half-primary-winding resistances in parallel.

The operation can be described by the equivalent circuit [Fig. 2(c)] that represents the double-pole change-over circuit.
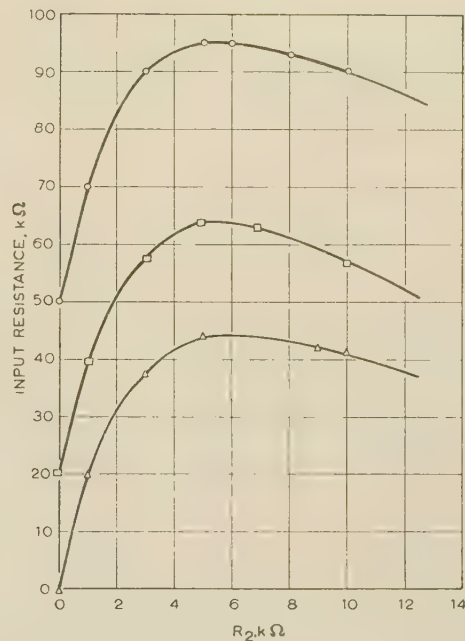
Fig. 6.—Input resistance as a function of stopper resistance for circuit 4 (high resistance).

—○—○— $R_1 = 50\,k\Omega$
—□—□— $R_1 = 20\,k\Omega$
—△—△— $R_1 = 0$

C_1.	Large	R_3'. 120 kΩ (core loss)	t_1. 20 millisec
L.	63 H	$R_{2(opt)}$. 5·25 kΩ (calculated)	t_2. 9·5 millisec
R_p.	120 Ω	k. 10	t_3. 0·25 millisec
R_s.	6·6 kΩ		t_4. Negligible

The two equivalent circuits are identical and the same expressions hold provided that the stopper resistance is denoted by $R_2 - R_p$, the half-primary resistance is denoted by $2R_p$ and we bear in mind that $t_4 = 0$.

(5.6) Effect of Capacitance Across Transformer

All transformers have some self-capacitance, and transformer-coupled chopper circuits are often seen with additional capacitance connected across a winding, usually of a value which causes resonance at the operating frequency. It is therefore interesting to analyse the effect of capacitance.

(5.6.1) Effect of Capacitance on Input Resistance.

Consider a capacitance C_T across the primary of the transformer in circuit 4, of value not greater than that required to cause resonance with L; i.e.

$$C_T \leqslant \frac{1}{\omega^2 L} = \frac{1}{4\pi^2} \frac{t_1^2}{L}. \quad . \quad . \quad . \quad (28)$$

If the waveform is not to have excessive droop, $R_2 \gg L/t_2$; hence

$$\frac{R_2 C_T}{t_2} \leqslant \frac{1}{4\pi^2} \left(\frac{t_1}{t_2}\right)^2 \quad . \quad . \quad . \quad . \quad (29)$$

In practice t_1/t_2 seldom exceeds about 3. Hence the time-constant $R_2 C_T \ll t_2$ and C_T charges rapidly at the start of every period t_2. Therefore it modifies the voltage across L only slightly, having at the end of t_2 a potential

$$V_1 \varepsilon^{-R_2 t_2/L} \quad . \quad . \quad . \quad . \quad . \quad (30)$$

When the switch reverses, the capacitor is short-circuited and its charge is lost. (Even if it is, in fact, a distributed capacitance, or one across the secondary, only the winding resistances are

involved, and the time-constant is much less than the short-circuit time, $2t_3$.)

The process is repeated twice every cycle, and the current drained from the filter capacitor is

$$2\frac{C_T V_1}{t_1}\varepsilon^{-R_2 t_2/L} \quad . \quad . \quad . \quad . \quad . \quad (31)$$

The effect of C_T is therefore equivalent to that of a resistor

$$R_c = \frac{1}{2}\frac{t_1}{C_T}\varepsilon^{+R_2 t_2/L} \quad . \quad . \quad . \quad . \quad (32)$$

connected across the filter capacitor, and the input resistance is reduced accordingly. Experimental checks have shown the above analysis to be accurate to within a few per cent.

Practical values of self-capacitance have a negligible effect even when the load is only the core-loss resistance. Capacitance sufficient to cause resonance with the winding, however, may lower the input resistance considerably, e.g. by about 20% in one case (see Section 6.1).

(5.6.2) Filtering Effect.

A high-frequency voltage injected with the direct input voltage is attenuated by R_1 and C_T. This effect may be of some use. Of greatest interest, however, is the effect on pick-up at the operating frequency ω.

A voltage of this frequency across C_1 is presented to the transformer substantially as a second and higher harmonics and a possible d.c. component. The impedance of the resonated transformer at the second-harmonic frequency is approximately

$$\frac{-j}{2\omega C_T} = \frac{-j\omega L}{2} \text{ (since } C_T = 1/\omega^2 L) = -j\pi L/t_1 \quad . \quad (33)$$

The double-frequency voltage is thus reduced, owing to R_2 and C_T, by an attenuation factor which is not less than

$$\frac{-j\pi L/t_1}{L/t_2 - j\pi L/t_1} = \frac{1}{1 - j\frac{1}{\pi}\frac{t_1}{t_2}} \quad . \quad . \quad . \quad (34)$$

whose scalar magnitude is

$$\frac{1}{\sqrt{\left[1 + \frac{1}{10}\left(\frac{t_1}{t_2}\right)^2\right]}} \quad . \quad . \quad . \quad . \quad (35)$$

If t_1/t_2 does not exceed about 3, the attenuation factor lies between $1\cdot0$ and $0\cdot7$.

Since the resonating capacitor lowers the input resistance of the circuit, it also lowers the desired output of chopped direct current. Thus the ratio between the chopped d.c. output and the unwanted output is scarcely improved by the presence of C_T, and may actually be made worse. This has been confirmed by experiment (Fig. 7). It may thus be concluded that additional capacitance connected across a winding in circuits with an overlap contact arrangement serves little useful purpose.

(5.6.3) Buffering Effect.

A small amount of capacitance is desirable, however, in order to enable the magnetizing current to continue flowing if brief open-circuits of the transformer occur, owing to faulty switching, for example. In the complete absence of capacitance the briefest open-circuit causes the flux to collapse, high transient voltages to be set up, a loss of energy and a considerable drop in input resistance. The self-capacitance of the winding provides sufficient buffering for brief open-circuits.
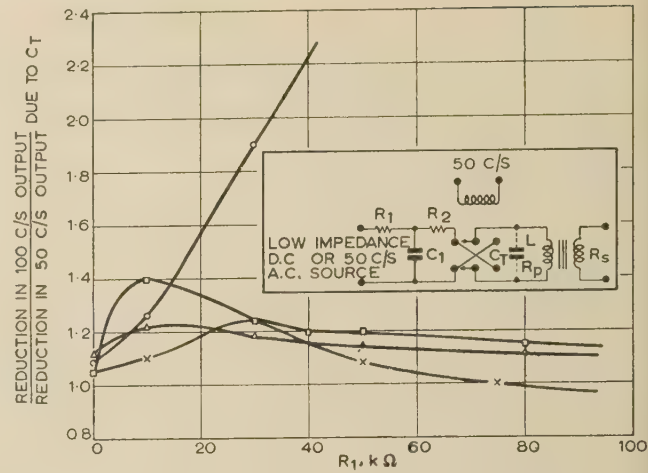
Fig. 7.—Filtering effect of capacitance across transformer (circuit 4)

—×—×— $C_1 = 4\,\mu$F
—△—△— $C_1 = 1\,\mu$F
—☐—☐— $C_1 = 0\cdot25\,\mu$F
—○—○— $C_1 = 0$

C_1, $0\cdot15\,\mu$F
R_2, $5\cdot25$ kΩ
Other components as in Fig. 6
A.C. signal across C_1 in phase with contacts.

(6) TRANSFORMER-COUPLED CIRCUITS: DISCUSSION

(6.1) Circuit Characteristics with Typical Components

Table 3 gives characteristics calculated for the double-pole change-over circuit used with typical components under high-resistance conditions.

The transformer was astatically wound on a Mumetal core and potted inside a double shield of the same material to eliminate induced pick-up effects.[9,12] The chopper had the contact arrangement shown in Fig. 4, contacts A and E being connected for change-over operation. This arrangement has the property that, theoretically, the output contacts B and D must always be

Table 3

CHARACTERISTICS OF CIRCUIT 4 WITH TYPICAL COMPONENTS AT 50 c/s

Transformer*

Turns ratio	10
Magnetizing inductance	45 henrys
Core-loss resistance (40–60 c/s)	100 kilohms
Self-capacitance	$0\cdot016\,\mu$F
Primary resistance	190 ohms
Secondary resistance	60 ohms

Chopper

t_1	20 millisec
t_2	As adjusted
t_3	$0\cdot16$ millisec

Circuit 4 (high resistance)

	$t_2 = 10$ millisec kilohms	$t_2 = 6\cdot67$ millisec kilohms
R_3'	100	100
L/t_2	$4\cdot5$	$6\cdot75$
$6\frac{t_1}{t_2}\left(\frac{L}{t_2}\right)^2\frac{1}{R_3'}$	$2\cdot4$	$8\cdot2$
$R_{2(opt)} = \frac{L}{t_2}\sqrt{\frac{24t_3}{t_2}}$ ..	$2\cdot8$	$5\cdot1$
Input resistance	$R_1 + 30$	$R_1 + 52$
Shunt resistance equivalent to C_T		
$R_c = \frac{1}{2}\frac{t_1}{C_T}\varepsilon^{R_2 t_2/L}$	1 200	1 300

* All values referred to primary.

connected either together or to the input contacts A and C. Moreover, the latter are short-circuited for only a small fraction of a cycle ($4t_3/t_1 = 3\%$ nominal and 5% maximum in this case) which is virtually independent of the contact setting.

The circuit is assumed to be terminated by a valve with no grid resistor, so that R_3' is equal to the core-loss resistance of the transformer referred to the primary.

The assumptions on which the formulae were based are evidently justified in this case. For instance, $R_2 < L/t_2$ and $R_2 \ll R_3'$. R_1 must exceed $2 \cdot 4$ or $8 \cdot 2$ kilohms; there would often be cases in which it did so with the input resistances quoted, so that this is a reasonable assumption to use in calculating R_2.

(6.2) Advantages of Reducing t_3

Every time the chopper reverses, the filter capacitor is short-circuited through R_2 and energy is wasted. It is interesting to examine the effects of reducing the duration of the short-circuit (t_3 per half-cycle) towards zero. In the type of change-over chopper shown in Fig. 4, t_3 is a function of the design, which would have to be modified. In the usual single-pole double-throw arrangement, t_3 depends on the contact setting and must be sufficiently large to allow for drifts in setting which could otherwise open-circuit the transformer during change-over. The difficulty can be avoided by using contacts ABC or CDE of the change-over type of chopper (Fig. 4). As t_3 approaches zero, so does the optimum value of R_2, and a very low impedance is maintained across the transformer throughout the cycle. This gives two practical advantages. First, it proportionately reduces the voltages developed across the transformer by stray currents which by-pass the inter-winding screen (Fig. 5): these currents can be a serious source of trouble when the circuit cannot be earthed on the primary side. Secondly, it proportionately reduces the effect of voltages induced in the transformer by stray fields, and so reduces the shielding requirements (the source impedance of these voltages is ωL or $k^2\omega L$, where ω is their frequency).

The other main effect of reducing t_3 is to raise the input resistance. Putting $t_3 = 0$, and assuming that R_3' is high (equal to the core loss resistance) in eqn. (10) we find that the upper limit of input resistance is

$$R_1 + \frac{R_3'}{\dfrac{2}{t_1}[t_2 + R_a R_3' t_2^3/12L^2]}$$

The second term in brackets becomes negligible as R_a diminishes towards its minimum value of R_p (circuit 4) or $2R_p$ (circuit 5). Hence the upper limit of input resistance when there is no load resistor is $R_1 + (t_1/2t_2)R_3'$, i.e. it is limited by the core-loss resistance of the transformer. The shunting effect of transformer self-capacitance is negligible at mains frequencies, but may be significant at higher frequencies.

Table 3 shows that the input resistance for these components could be raised from $R_1 + 30$ kilohms to $R_1 + 100$ kilohms ($t_1 = 10$ millisec) or from $R_1 + 50$ kilohms to $R_1 + 150$ kilohms ($t_1 = 6 \cdot 67$ millisec), if t_3 were reduced to zero. This would be useful: more useful still would be the reduction of R_2 and the effect of pick-up.

(6.3) Comparison of Circuits 4 and 5

The waveform given by circuit 4 can be adjusted to have long or short zero steps during switching without loss of efficiency. This does not apply to circuit 5, in which the filter capacitor is short-circuited for the full duration of the step. (If the signal is passed through an RC-coupled amplifier and allowed to limit, a phase advance is obtainable which depends on the size of the

steps: this can be useful.) Circuit 4 permits simpler transformer windings but requires a more complicated chopper with four pairs of contacts, which may be less reliable.

The optimum stopper resistance is lower by an amount R_p for circuit 5 than for circuit 4. In very-low-resistance circuits the optimum may be negative, in which case circuit 4 has a slight advantage.

(6.4) Comparison with Underlap Circuits

Circuits like 4 and 5, but with underlap contact arrangements, no stopper resistor, and a buffer capacitor to absorb the magnetizing current during change-over, are used in vibrator power supplies. They are fully analysed by Evans,[5] and, with ideal buffering, can provide an input resistance approaching that of circuits 4 and 5 with zero overlap. This is a clear advantage. They also provide a flat-topped waveform. However, they suffer from the following disadvantages which may be serious in low-level circuits:

(a) Ideal buffering requires a capacitance related to L, R_3 and the underlap time. In order to allow for drifts in contact setting and manufacturing variations in L, over-buffering is common practice, although this lowers the input resistance.

(b) The output-voltage waveform is approximately trapezoidal, and there are no zero steps during switching.

(c) The output impedance is very high during the underlap period. This permits sudden increases in the effects of stray currents caused by external pick-up [Fig. 5(a)], valve hum, pick-up in the transformer, etc. The transients thus set up may block the following amplifier or cause errors in measurement.

(6.5) Factors in Transformer Design

The transformer characteristics which set an upper limit to the input resistance are the inductance and the core-loss resistance. Both of these can be increased by winding more turns, and the limit is set by space and the gauge of the wire required for the secondary. For very-low-resistance working the greatest efficiency occurs when $R_s' + R_p$ is a minimum. This requires that R_s' and R_p should be approximately equal. In practice it may be more useful to allot the secondary more than half the volume and so raise the inductance at the cost of increased primary resistance.

(6.6) Types of Filter Capacitor

Capacitance-coupled circuits and the high-resistance transformer-coupled circuits operate with filter capacitances of a few microfarads or less. The dielectrics may be of paper, metallized paper or polystyrene.

The problem is more difficult when low-resistance circuits call for capacitances of tens or hundreds of microfarads. Ordinary or non-polarized electrolytic capacitors can then be used with success if the measured voltages are not too small. Errors of 10–$20\,\mu\text{V}$ are commonly found, owing to currents generated within the capacitors, which act as batteries of high internal resistance. Tantalum electrolytic capacitors show promise of being satisfactory at low voltage levels, owing to their very high internal resistance.

(8) REFERENCES

(1) AUMONT, R., and ROMAND, J.: 'Dispositif électronique pour le mésure précis de tensions continues par comparaison', *Revue Générale d'Électricité*, 1953, **62**, p. 210.

(2) DAUPHINEE, T. M.: 'An Isolating Potential Comparator', *Canadian Journal of Physics*, 1953, **31**, p. 577.

(3) DAUPHINEE, T. M.: 'Method of Reducing Zero Error and Drift in Breaker Type D.C. Amplifiers', *Review of Scientific Instruments*, 1955, **26**, p. 401.

(4) DAUPHINEE, T. M., and WOODS, S. B.: 'Low-Level Thermocouple Amplifier and a Temperature Regulation System', *ibid.*, p. 693.

(5) EVANS, R. H.: 'Vibratory Power Convertors', *Proceedings I.E.E.*, Monograph No. 109 R, September, 1954 (**102 C**, p. 62).

(6) GIACOLETTO, L. J.: 'Parallel-Tuned Circuit Periodically Switched to a D.C. Source', *RCA Review*, 1952, **13**, p. 386.

(7) KROEBEL, W.: 'Ein hochempfindlicher Gleichspannungsverstärker mit hohem Eingangwiderstand', *Zeitschrift für Physik*, 1952, **133**, p. 30.

(8) KROEBEL, W: 'A Method for the Amplification of Extremely Small Thermo-Electric Voltages', *Zeitschrift für Angewandte Physik*, 1953, **5**, p. 286.

(9) WILLIAMS, A. J., TARPLEY, R. E., and CLARK, W. R.: 'D.C Amplifier Stabilised for Zero and Gain', *Transactions of the American I.E.E.*, 1948, **67**, p. 47.

(10) MCADAM, W., TARPLEY, R. E., WILLIAMS, A. J.: 'D.C Amplifier with Reduced Zero Offset', *Electronics*, August, 1951, **24**, p. 128.

(11) WILLIAMS, A. J., CLARK, W. R., and TARPLEY, R. E.: 'Electronically Balanced Recorder for Flight Testing and Spectroscopy', *Transactions of the American I.E.E.*, 1946 **65**, p. 205.

(12) GUSTAFSON, W. G.: 'Magnetic Shielding of Transformers a Audio Frequencies', *Bell System Technical Journal*, 1938 **17**, p. 416.

(13) 'Improvements in Vibratory Type of Synchronous Converter', British Patent No. 647551.

(14) 'Apparatus for Measuring, Recording or Controlling pH' British Patent No. 695909, May, 1951.

(15) 'The Measuring of Variables', British Patent Application No. 28566, 1954.

THE MEASUREMENT AND PREDICTION OF INDUCTION MOTOR STRAY LOSS AT LARGE SLIPS

By T. H. BARTON, Ph.D., Associate Member, and V. AHMAD, M.Sc., Graduate.

SUMMARY

It is shown that the stray loss in induction motors, at fractional slips greater than about 0·25, is almost entirely due to tooth ripple in the zigzag leakage flux. The stray loss is large; it can double the torque at a slip of 2·0 and can lead to serious errors when calculating accelerating and braking performance. In the normal operating range, a slip less than 0·05, although the stray loss is much lower it is still appreciable, and the component associated with tooth ripple in the main flux is of importance. Experimental evidence, both from actual machines and from high-frequency loss in samples of laminations, indicates that the zigzag leakage flux component of the loss is approximately proportional to (speed)$^{1·6}$ and (current)$^{1·9}$, and the results of the American Standard reverse-rotation test extrapolated according to this law are sufficient, together with the no-load stray loss, to predict the total stray loss over the whole practicable speed range.

A simple method for calculating the stray loss is given and its limitations are indicated. Calculated values of the loss are compared with the measured losses, but no definite conclusions as to the accuracy of the calculations can be drawn until more is known of the high-frequency loss in machine laminations.

LIST OF SYMBOLS

B = Flux density, gauss.
B_m = Mean flux density over a pole pitch, gauss.
D_f = Hysteresis factor.
f = Frequency, c/s.
I = Primary phase current, amp.
K = Stray-loss constant.
m = Lamination thickness, cm.
N_a = Number of slots.
n = Speed of rotation, r.p.m.
n_0 = Synchronous speed, r.p.m.
P = Power.
P_e = Eddy-current component of surface loss, watts/cm^2.
P_h = Hysteresis component of surface loss, watts/cm^2.
P_r = Rated output power, watts.
P_{st} = Stray loss, watts.
s = Slip.
T = Torque.
V_1 = Primary phase voltage.
V_0 = Primary e.m.f. per phase, volts.

$\left.\begin{array}{c} \\ \\ \end{array}\right\}$ = Stray loss factors.

$\left.\begin{array}{c} \\ \\ \end{array}\right\}$ = Flux oscillation factors.

λ = Slot pitch, cm.
η_1 = Hysteresis factor.
Φ = Flux, maxwells.
θ = Absolute temperature of winding.

(1) INTRODUCTION

Losses in addition to the normal copper, iron and friction losses occur in all rotating electric machines. They are usually termed 'stray losses' and are principally due to high-order flux harmonics produced by the winding slots. In induction motors, the combination of extremely short air-gaps with slotting in both stator and rotor increases these harmonics so much that the magnitude of the stray loss is relatively far greater than in direct-current and synchronous machines. A typical induction motor torque/speed characteristic, together with that obtained from the normal induction action, is shown in Fig. 1. It is to be noted that the retarding torque due to stray losses, at a slip of 2·0, is of the same order of magnitude as the induction torque.

Although the discrepancy between expected and actual torques is most marked, it does not appear to be widely appreciated. This may be accounted for by two facts: first, the stray loss is small at small slips covering the normal operating range; secondly, any discrepancy between measured and predicted torque in this region is more readily attributed to a very much smaller error in speed measurement. However, the present tendency towards better utilization of material and operation at high slips for acceleration, reverse-current-braking and asymmetrical voltage control makes a better understanding of this loss most desirable.

Much theoretical and practical work on stray loss at small slips has been published, the latest contribution being that of Rawcliffe and Menon.[1] The general conclusion is that this loss is small in magnitude, being of the same order as the line-frequency iron loss. The only published data regarding stray loss at large slips are those derived by Morgan, Brown and Schumer[2,3] and extended by Ware.[4] Their results, while giving no theoretical explanation of the loss, are most valuable since they indicate that it is a function of the square of the current and the square of the speed of rotation. This has resulted in the formulation of a simple test for the measurement of stray loss, the reverse-rotation test, which is now incorporated in the induction motor test code of the American Institute of Electrical Engineers.[5] The measurements made by the authors suggest that the index of variation with speed should be reduced from 2·0 to about 1·6, and that the results of the reverse-rotation test are not directly applicable to operation at small slips, allowance having to be made for the no-load stray loss. This is supported by the published data for Lohys sheet steel[6] and by Dannat's measurements[7] of loss in Stalloy at medium audio frequencies.

The discovery that the stray loss depends upon the square of current, rather than of voltage, is of the greatest importance. It suggests that the source of loss is the leakage flux, rather than the main flux as has been previously supposed, and it is claimed that the calculations of Section 8 confirm this view.

The present paper is the result of an experimental study of the stray losses of four medium-sized induction motors over a

wide speed range. As indicated earlier, the results differ in detail from those published by Morgan. The experimental method is also different from Morgan's because of the necessity to consider harmonic torques in two of the machines. An analysis of the source and magnitude of loss is given which is simple, but probably sufficiently accurate for practical purposes. The major source of stray loss is shown to be slot ripple in the zigzag leakage flux and, by comparison, the loss due to ripple in the main flux is shown to be small.

(2) EXPERIMENTAL PROCEDURE

The stray losses of the four machines listed in Table 1 were measured over the slip range 0–2·0 by the following method. The electrical characteristics of each machine were measured over the whole speed range at four different voltages, and their exact equivalent circuits were determined in the manner described

Table 1

Details of the Machines Tested

Machine number	Horse-power	Synchronous speed, r.p.m.	Type of secondary
1	7·5	1 500	Squirrel-cage
2	10	1 500	Squirrel-cage
3	15	1 000	3-phase
4	20	1 500	3-phase

Table 2

Experimentally Obtained Values of α, β and K in Eqn. (1). n_0 R.P.M. is the Synchronous Speed

Machine number	Speed range	α	β	K
1	$0—-n_0$	1·51	1·90	$3·3 \times 10^{-3}$
	$0-+n_0$	1·51	1·90	
2	$0—-n_0$	1·52	2·04	$4·7 \times 10^{-3}$
	$0-+n_0$	1·08	2·00	
3	$0—-n_0$	1·51	1·90	$6·4 \times 10^{-3}$
	$0-+n_0$	1·20	1·51	
4	$0—-n_0$	1·49	1·89	$11·7 \times 10^{-3}$
	$0-+n_0$	1·46	1·87	

by Morris.[8] Their output torques due to the induction motor action were calculated from these circuits and were compared with the actual torques as measured by a calibrated d.c. machine. This machine was operated at a low flux density, so that flux loss due to armature reaction and increase in stray loss with load current were small. Because of this, the majority of results needed no correction for these effects. The difference between the predicted and the measured torque is the retarding torque due to stray losses, and from this the stray-loss power is derived.

Numerous refinements of the above procedure were necessary in order to obtain consistent results. Although skin effect in the rotor bars of the two squirrel-cage machines was negligible over the frequency range considered, allowance had to be made in all machines for the increase in winding resistance with increase in temperature. The stator-winding temperature was measured by the resistance method immediately after each experimental result, and sufficient time was taken between readings to keep the temperature variations within reasonable limits. The rotor-winding temperature was assumed equal to that of the stator winding, a reasonable assumption for all machines except No. 2, whose rotor resistance is about three times the normal value. However, the absence of a method of

measuring rotor temperature during the tests necessitated th assumption in this case also.

Allowance was also made for the variation of leakage reactan with current. The leakage reactances of machines Nos. 1 and decrease slowly with increasing current for currents exceedin 1·2 per unit. Owing to its rotor slots being closed, the lea age reactance of machine No. 3 is large at small currents b decreases rapidly with increasing current to a more-or-less stead value. Short-circuit characteristics were used to establish th current/leakage-reactance relationship in all cases.

The fifth-harmonic torques of machines Nos. 1 and 2 wer calculated from experimentally derived fifth-harmonic para meters. The seventh harmonics, although present, were to small to be determined by this method and were ignored. Th resultant errors will be small, probably no greater than th experimental error, and will occur at low speeds when the stra loss is unimportant. It is to be noted here that the method o torque measurement used by Morgan, in which the powe crossing the air-gap is equated to the torque in synchronou watts, is vitiated by the presence of harmonics.

(3) DISCUSSION OF THE RESULTS

The measured and predicted torque/speed characteristics o machine No. 1 are shown in Fig. 1. The discrepancy betwee
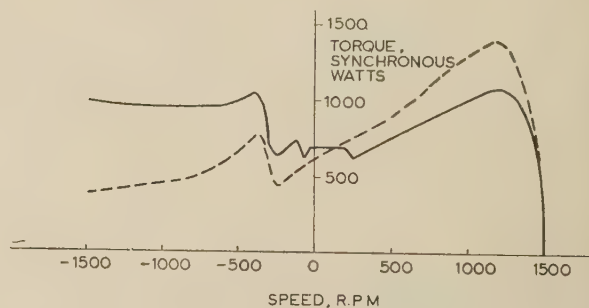
Fig. 1.—Torque/speed characteristics for machine No. 1 at 119 vol per phase.
——— Measured.
- - - - - Calculated.

the two is large, much larger than is generally considered to b the case, but is typical of all induction motors within the author experience.

The stray-loss power was plotted, on log-log paper, as function of current at constant speed and speed at consta current, the results for machine No. 4 being given in Fig. Within reasonable limits, all the graphs were linear, indicatin that the relationship between stray loss P_{st}, current I, an speed n, may be written as

$$P_{st} = Kn^{\alpha}I^{\beta} \quad \ldots \quad \ldots \quad ($$

where K, α and β are constants. The values of the constan are given in Table 2, from which it is seen that

$$P_{st} = Kn^{1·6}I^{1·9} \quad \ldots \quad \ldots \quad ($$

is a close approximation to the measured results in all cas except the positive speed ranges of machines Nos. 2 and 3. may be that in this speed range the stray loss does follo different law in these two machines, but it is considered mo probable that the difference is due to experimental error. both cases the stray loss at positive speeds is obtained as the ve small difference of two large quantities with a conseque increased possibility of error in the results.

Allowing for this point, the coincidence of the positive a
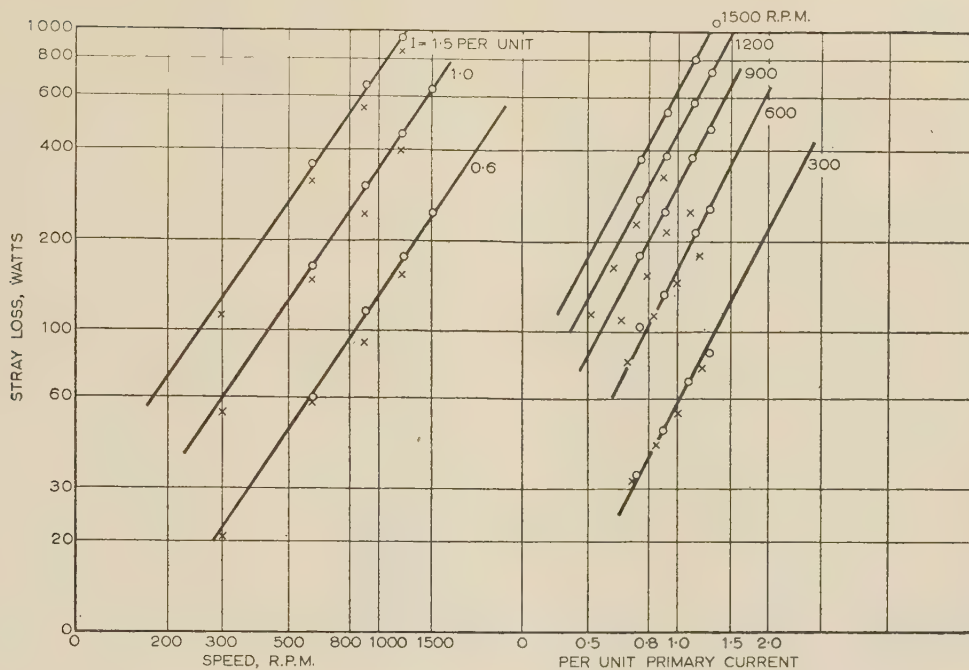
Fig. 2.—Stray-loss characteristics of machine No. 4.

○ Negative speed experimental points.
× Positive speed experimental points.

gative speed/loss characteristics at constant current supports the view that current rather than voltage is the parameter of importance when considering stray loss, since the voltages at constant current and positive and negative speeds are widely different. This was checked by measuring the stray loss of machines Nos. 1 and 2 at synchronous speed and $1/\sqrt{3}$ times the rated voltage. The losses were found to be 18 watts and 43 watts, respectively. The voltages used for measurements at high slips did not exceed one-half the rated value, and it is concluded that the stray loss associated with the main flux is small compared with that associated with the leakage flux, except at very small slips.

With regard to the magnitude of the indices α and β, it is noteworthy that the high-frequency loss in laminations of the same thickness and of similar material to those used in machines Nos. 1 and 2 follows closely the law[6]

$$P = 128 \times 10^{-12} f^{1 \cdot 6} B^{1 \cdot 9} \text{ watts/kg} \quad . \quad . \quad (3)$$

over the range 400–2 000 c/s, 100–2 000 gauss.

(4) THE PREDICTION OF STRAY LOSS

A rigorous analysis of the problem of induction motor stray loss is extremely complicated and requires data which are not available; e.g. the iron loss due to three simultaneously applied fluxes of different frequencies and the variation of flux density above semi-closed slots when the tooth tips are saturated. Hence, in order to present a compact overall view of the problem, the method of calculation described is highly simplified on points of detail. Nevertheless it probably yields results of sufficient accuracy for most purposes. The simplifying assumptions and points where present knowledge is inadequate are indicated.

Calculation of stray loss has been carried out only for machines Nos. 1 and 2, since design details were not available for machines Nos. 3 and 4. Details of the calculations for machine No. 1 are given in an Appendix, Section 8.

Since the tooth ripple associated with the main flux is generally much smaller than that in the zigzag flux, it will be assumed that the two phenomena can be analysed independently.

(4.1) Loss Due to Tooth Ripple in the Main Flux

The loss due to tooth ripple in the main flux has three components: the first is due to flux-density variations in the air-gap surfaces of the laminations, the second to pulsations in the tooth fluxes, and the third to high-frequency circulating currents in a squirrel-cage winding. The latter component has the effect of damping the flux pulsations in the rotor teeth and correspondingly increasing the pulsations in the stator teeth. A phase winding has such a small response to the high-order ripple harmonics that it may be neglected, and, fortunately, the rotor bars of a squirrel-cage are usually skewed by one stator slot pitch, i.e. by one cycle of ripple, and are effectively open-circuited. In this latter case, some circulation can be set up if the bars are in contact with the laminations, but, owing to the high impedance of the latter, it can be neglected.

(4.1.1) The Surface Loss.

This has been investigated by Gibbs[9]; his expressions for the eddy-current loss, P_e, and the hysteresis loss, P_h, in watts per square centimetre, are given in eqns. (4) and (5).

$$P_e = 2 \cdot 14 m \lambda \beta_2^2 B_m^2 (n N_a)^{\frac{3}{2}} \times 10^{-4} \quad . \quad . \quad . \quad (4)$$

$$P_h = 6 \cdot 67 \eta_1 D_f (n N_a)^{\frac{1}{2}} (\beta_1 B_m)^{1 \cdot 6} \times 10^{-9} \quad . \quad . \quad (5)$$

Values of β_1, β_2, η_1, and D_f are taken from Gibbs's paper.

Although the above expressions were derived assuming an unslotted pole-face, they have been applied directly to machines Nos. 1 and 2, since these have semi-closed slots and short air-gaps, thus minimizing the interference of the stator and rotor slot effects.

(4.1.2) Loss Due to Tooth Flux Pulsation.

It is evident that the tooth flux, apart from its supply frequency

alternation, varies between a maximum value when a tooth is opposite a tooth and a minimum when a slot is opposite a tooth. Excepting the small effects of interference between stator and rotor slots, the variation of flux density over, say, a rotor tooth due to a stator slot can be determined by the use of Carter's fringing curves. The tooth flux, and its variation with relative stator and rotor position, is then readily calculated, and Hoseason[10] has shown that the results so obtained are reasonably close to those obtained in practice. However, in a modern machine, with semi-closed slots, there is usually some degree of saturation of the tooth tips when loaded, so that the effective slot opening is greater than the actual slot opening. Because of uncertainty as to the effective slot opening, the method of calculation of tooth flux pulsation given by Chapman[11] has been adopted. Briefly, this assumes a constant air-gap flux density between tooth and tooth, and zero flux density over a slot opening. With unsaturated tooth tips, Hoseason's measurements show that this method gives values of pulsation considerably higher than those obtained in practice, and it is therefore probable that the calculated values of tooth pulsation loss are too great.

Having determined the variation of tooth flux density, its fundamental component has been derived by harmonic analysis and the iron loss determined by using eqn. (3).

(4.2) Loss Due to Tooth Ripple in the Zigzag Flux

Fig. 3 shows the two zigzag leakage fluxes, Φ_1 and Φ_2, in a tooth due to the slot currents on either side. It is well known
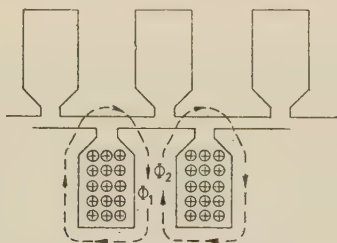
Fig. 3.—Tooth flux produced by zigzag leakage.

that Φ_1 and Φ_2 vary with the relative positions of stator and rotor, the derivation of the value as a function of relative position appearing in the majority of books on machine design. However, it is not widely appreciated, and in fact the authors have found only one relevant reference,[12] that, unless the stator and rotor slot pitches are equal, the variations of Φ_1 and Φ_2 are not in phase and there is a resultant variation of tooth flux at the tooth ripple frequency.

The method of calculation adopted is to determine the zigzag flux due to the current in a slot as a function of relative rotor position by the well-known method, i.e. that given by Vickers.[13] Its fundamental component is then obtained and the variation of tooth flux density is derived. The tooth loss is then obtained from eqn. (3).

When the slot currents on either side of the tooth are in phase, the resultant tooth flux is obtained simply as the difference of the components Φ_1 and Φ_2. However, if the currents are not in phase, as, for example, for squirrel-cage rotor teeth or for teeth between phase belts, the superposition of the supply frequency and the slot ripple frequency results in the resultant tooth flux having three components. One of these alternates at the supply frequency, another at a frequency equal to the difference between the tooth ripple frequency and the supply frequency, and the third at a frequency equal to the sum of the tooth ripple frequency and the supply frequency. Iron loss values under such conditions are not available, and since the

slot ripple frequency is normally very much greater than the supply frequency, the vector mean of the two slot currents is taken as the equivalent in-phase current. Since iron loss is not linearly dependent on the amplitude of flux density variation, it is expected that the beats produced by the two different frequencies noted above will result in a loss somewhat higher than that obtained by the simplified calculation.

(4.3) Comparison of the Calculated and Measured Values

The calculated components of the tooth ripple loss and the experimentally derived equations for machines Nos. 1 and 2 are given in Table 3. The equations are rewritten in terms of per-unit speed ($n/1\,500$), per-unit current (I/I_{fl}) and per-unit voltage ($V_0/400$).

When interpreting these equations it must be remembered that V_0 is the e.m.f. of the machine and therefore decreases as the slip increases, being approximately half the applied voltage at standstill, owing to the voltage drop of the primary leakage reactance. However, when operating at normal voltages and high slips, saturation of the tooth tips results in increased effective slot opening with a consequent increase in loss.

The calculated results indicate that, for fractional slip exceeding $0 \cdot 05$, the stray loss associated with the zigzag flux predominates over that associated with the main flux; e.g. at a slip of $0 \cdot 25$ the former loss is about five times the latter in both machines. In the normal operating range, below $0 \cdot 05$ slip, the two components of the stray loss are of more-or-less equal importance.

In both cases the calculated loss is considerably lower than the measured loss. This error is to be expected from a consideration of the main sources of error in the calculations:

(a) The iron loss formula has been extrapolated to flux densities higher than those for which experimental results are available. Any error due to this is thought to be small.

(b) The iron loss equation is derived from measurements on a carefully constructed core, and the actual lamination loss will be considerably higher than that given by the equation. Clayton,[14] when referring to low-frequency loss, states that the eddy-current loss in the teeth of a carefully made machine is normally of the order of three times the loss in a carefully made-up sample of the laminations.

(c) The tooth pulsations in the main flux and zigzag flux have been treated separately. Since they are of the same frequency it would be more accurate to consider the resultant combined pulsation.

Where, as is usually the case, it is desirable to minimize the stray loss, the ratio of primary to secondary slots should be as near unity as possible, and the number of slots per pole should be as small as possible. This will reduce both the magnitude and frequency of the flux ripples, but is in conflict with the requirements for freedom from rotor locking, crawling and objectionable noise, so that some compromise is necessary in practice.

(5) CONCLUSIONS

The great effect of stray loss on induction motor performance, and the fact that this loss is a function of current rather than voltage, noted by Morgan, Brown and Schumer,[2,3] are confirmed. However, the experimental results described here indicate that the indices of variation of the loss with speed and current should be about $1 \cdot 6$ and $1 \cdot 9$, respectively, rather than $2 \cdot 0$ in each case as proposed by Morgan et al. This conclusion is supported by the published data regarding iron loss under the relevant conditions.

The method, used by Morgan, for determining the torque produced by the normal induction action is shown to be inaccurate when the torques produced by space harmonics of flux are appreciable.

Theoretical considerations indicate that the principal source of loss at slips greater than about 0·1 is due to high-frequency variations in the component of tooth flux produced by the zigzag leakage. These variations could be eliminated by making the rotor and stator slot pitches equal, but this is not practicable owing to rotor locking and crawling effects. However, the variations can be minimized by having the two pitches as nearly equal as possible and by using the least possible number of slots per pole. This is a well-known rule of induction motor design[12] adopted in order to reduce high-frequency variations in the tooth component of the main flux.

Because of its dependence on approximately the square of the current, the stray loss associated with zigzag leakage flux is much reduced under normal operating conditions, i.e. below 0·05 slip. However, it is by no means negligible, and in combination with the loss associated with the main flux, can be an appreciable proportion of the motor output, especially when it is considered that it is supplied via the mechanical output of the machine, thus entailing extra power-frequency loss. As an example, the stray loss of machine No. 1, at full load and normal voltage, is about 300 watts, i.e. about 5% of its output. Under these conditions the stray losses associated with the main flux and the leakage flux are comparable, so that neither the reverse rotation test nor the determination of the no-load stray loss is individually sufficient to determine the loss. It is essential to combine the results of the two tests in the manner indicated by the expressions of Table 3.

(6) ACKNOWLEDGMENTS

The authors gratefully acknowledge their indebtedness to the University of Sheffield for the experimental facilities afforded them, and to Mr. O. I. Butler for his advice and encouragement.

(7) REFERENCES

(1) RAWCLIFFE, G. H., and MENON, A. M.: 'A Simple New Test for Harmonic-Frequency Losses in A.C. Machines', *Proceedings I.E.E.*, Paper No. 1222 U, April, 1952 (**99**, Part II, p. 145).
(2) MORGAN, T., BROWN, W. E., and SCHUMER, A. J.: 'Reverse Rotational Tests for the Determination of Stray-Load Loss in Induction Motors', *Transactions of the American Institution of Electrical Engineers*, 1939, **58**, p. 319.
(3) MORGAN, T., BROWN, W. E., and SCHUMER, A. J.: 'Induction Motor Characteristics at High Slip', *ibid.*, 1940, **59**, p. 464.
(4) WARE, D. H.: 'Measurement of Stray-Load Loss in Induction Motors', *ibid.*, 1945, **64**, p. 194.
(5) 'American Standard Test Code for Polyphase Induction Motors and Generators' (American Standards Association, New York, 1954).
(6) Sankey Data Book: 'High Frequency Losses in Lohys'.
(7) DANNAT, C.: 'The Variation of the Magnetic Properties of Ferromagnetic Laminae with Frequency', *Journal I.E.E.*, 1936, **79**, p. 667.
(8) MORRIS, D.: 'Some Tests of an Exact Theory of the Transformer', *Proceedings I.E.E.*, Paper No. 902 S, February, 1950 (**97**, Part II, p. 17).
MORRIS, D.: 'Some Tests of an Exact Practical Theory of the Induction Motor', *ibid.*, Paper No. 1041 U, December, 1950 (**97**, Part II, p. 767).
(9) GIBBS, W. J.: 'Tooth Ripple Losses in Unwound Pole-Shoes', *Journal I.E.E.*, 1947, **94**, Part II, p. 2.
(10) HOSEASON, D. B.: 'Tooth Frequency Iron Losses in Slip Ring Induction Motors', *Electrician*, 1923, **91**, p. 240.
(11) CHAPMAN, F. T.: 'A Study of the Induction Motor' (Chapman and Hall, London, 1930), p. 258.
(12) ALGER, P. L.: 'The Nature of Polyphase Induction Machines' (John Wiley and Sons, Inc., New York, 1951), p. 174.
(13) VICKERS, H.: 'The Induction Motor' (Sir Isaac Pitman and Sons Ltd., London, 1953), p. 412.
(14) CLAYTON, A. E.: 'The Performance and Design of Direct Current Machines' (Sir Isaac Pitman and Sons Ltd., London, 1938), p. 120.

(8) APPENDIX: CALCULATION OF THE STRAY LOSS IN MACHINE No. 1

Machine No. 1 has a 3-phase 4-pole single-layer primary winding rated at 400 volts 6·3 amp per phase. There are 552 conductors in series per phase and the core length is 11·5 cm. Other dimensions are given in Fig. 4.

Table 3

STRAY-LOSS EXPRESSIONS FOR MACHINES NOS. 1 AND 2. ALL LOSSES ARE IN WATTS

	Machine No. 1	Machine No. 2
Surface loss due to main flux	$42(n/1\,500)^{1·5}(V_0/400)^{2·0}$	$44(n/1\,500)^{1·5}(V_0/400)^{2·0}$
Tooth pulsation loss due to main flux ..	$12·5(n/1\,500)^{1·6}(V_0/400)^{1·9}$	$118(n/1\,500)^{1·6}(V_0/400)^{1·9}$
Tooth pulsation loss due to zigzag leakage flux ..	$97(n/1\,500)^{1·6}(I/6·3)^{1·9}$	$244(n/1\,500)^{1·6}(I/9·0)^{1·9}$
Experimentally derived loss equation..	$206(n/1\,500)^{1·51}(I/6·3)^{1·9}$	$318(n/1\,500)^{1·52}(I/9·0)^{2·0}$

Fig. 4.—Slot dimensions, in centimetres for machine No. 1.

(8.1) Loss Due to Tooth Ripple in the Main Flux

The mean air-gap flux density, B_m, is related to the phase e.m.f., V_0 volts, by

$$B_m = 12 \cdot 5 \, V_0 \text{ gauss}$$

(8.1.1) The Surface Loss.

The hysteresis component of the surface loss is negligible, and by substitution in eqn. (4) the eddy-current component is

$$42(n/1\,500)^{1 \cdot 5}(V_0/400)^2 \text{ watts}$$

(8.1.2) Tooth Pulsation Loss.

Assuming zero flux density over slot openings, the mean air-gap flux density between teeth is $33 \cdot 6 \, V_0$ gauss. Hence the flux density in rotor tooth, B_{RT}, is

$$33 \cdot 6 \, V_0 \, \frac{\text{Breadth of rotor tooth top opposite stator teeth}}{\text{Breadth of rotor tooth} \times \text{core stacking factor}}$$

Taking a mean rotor-tooth breadth of $0 \cdot 6$ cm and a stacking factor of $0 \cdot 95$, the variation of the ratio B_{RT}/V_0 with rotor position is shown in Fig. 5, together with its fundamental component.
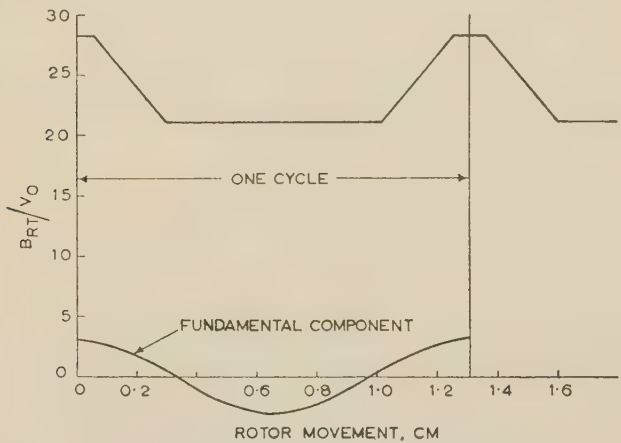
Fig. 5.—The variation of the main flux component of the rotor-tooth flux density with rotor position.

The variation of stator-tooth flux density is determined in a similar manner and is found to be negligible.

The use of eqn. (3) gives the rotor-tooth pulsation loss as

$$12 \cdot 5 \, (n/1\,500)^{1 \cdot 6}(V_0/400)^{1 \cdot 9} \text{ watts}$$

(8.2) Loss Due to Tooth Ripple in the Zigzag Leakage Flux

The zigzag leakage flux due to the current in one slot is calculated, and the flux density produced by it in adjacent teeth is then found, taking a mean breadth for the tapered teeth. The latter density varies with rotor position, and the amplitude of its fundamental component, B_s^l and B_r^l, in stator and rotor respectively, is found by harmonic analysis. In this case

$$B_s^l = 210 \text{ gauss/amp of stator phase current}$$
$$B_r^l = 290 \text{ gauss/amp of stator phase current}$$

The flux variations, due to in-phase currents in adjacent slots, have a phase difference of 84° in the stator teeth and 66° in the rotor teeth. Hence, the amplitudes, B_s and B_r, of the resultant tooth flux density variations, when the slot currents on either side of the tooth are in phase, are given by

$$B_s = 281 \text{ gauss/amp of stator phase current}$$
$$B_r = 316 \text{ gauss/amp of stator phase current}$$

On the stator, since there are three slots per pole and phase, only two-thirds of the teeth have the amplitude B_s. In the remaining third, since the slot currents on either side differ in phase by 60°, the amplitude is taken as $B_s \cos 30°$, i.e. 243 gauss/amp. On the rotor, since it has a squirrel-cage winding with 11 slots per pole, the phase displacement between currents in adjacent slots is $16 \cdot 35°$, and the amplitude of the tooth flux density ripple is taken as $B_r \cos 8 \cdot 2°$, i.e. 313 gauss/amp.

The frequency of the flux ripple is $0 \cdot 735n$ in the stator teeth and $0 \cdot 6n$ in the rotor teeth, and the loss produced by the ripple is obtained by substitution of the above values in eqn. (3).

$$\begin{aligned} \text{Total loss} &= 128 \times 10^{-12}[3 \cdot 33(0 \cdot 74n)^{1 \cdot 6}(281 \, I)^{1 \cdot 9} \\ &\quad + 1 \cdot 67(0 \cdot 74n)^{1 \cdot 6}(243 \, I)^{1 \cdot 9} \\ &\quad + 2 \cdot 45(0 \cdot 6n)^{1 \cdot 6}(313 \, I)^{1 \cdot 9}] \\ &= 97(n/1\,500)^{1 \cdot 6}(I/6 \cdot 3)^{1 \cdot 9} \text{ watts} \end{aligned}$$

The fraction $I/6 \cdot 3$ is used in the above expression since $6 \cdot 3$ amp represents the full load phase current.

(8.3) The Total Tooth Ripple Loss

Adding the various components, the expression for the total tooth ripple loss, P_{st}, becomes

$$\begin{aligned} P_{st} &= 42(n/1\,500)^{1 \cdot 5}(V_0/400)^2 + 12 \cdot 5(n/1\,500)^{1 \cdot 6}(V_0/400)^{1 \cdot 9} \\ &\quad + 97(n/1\,500)^{1 \cdot 6}(I/6 \cdot 3)^{1 \cdot 9} \text{ watts} \end{aligned}$$

THEORY AND EQUIVALENT CIRCUITS OF THE STATOR-FED POLYPHASE SHUNT COMMUTATOR MOTOR

By C. S. JHA, B.Sc., Graduate.

SUMMARY

Tensor analysis is used to develop a comprehensive theory of the stator-fed 3-phase shunt commutator motor. Before proceeding with the general analysis, the simple case of a machine where all the phase shift between the secondary standstill induced voltage and the secondary impressed voltage is either in the motor or the regulator is considered. From the equations of performance of the simple machine, the existing equivalent circuit proposed by Coulthard is derived. The limitations of this circuit are discussed and better circuits are proposed. For the general analysis, the machine has an auxiliary winding on the rotor displaced from the main winding for power-factor improvement, shift of the brushes from the neutral position, and a biased double regulator with completely dissimilar-component single regulators for speed and power-factor control. An equivalent circuit for the general machine is proposed and a method of measuring its parameters is indicated. Certain modifications to the theory to overcome inherent limitations are discussed.

(1) INTRODUCTION

The stator-fed polyphase shunt commutator motor is essentially an induction motor with auxiliary apparatus for injecting proper electromotive force at slip frequency into the rotor circuit for speed regulation and power-factor control. It thus consists of the following:

   (*a*) An induction motor with wound rotor, the rotor windings being connected to a commutator instead of being short-circuited.

   (*b*) A device to obtain a variable e.m.f. at supply frequency in phase with the rotor standstill e.m.f. for speed regulation.

   (*c*) A device to obtain an e.m.f., variable or fixed, at supply frequency in quadrature with the rotor standstill e.m.f. for power-factor improvement.

   (*d*) The commutator and brushes, which act as a frequency convertor for changing the injected e.m.f.'s of (*b*) and (*c*) from supply to slip frequency.

There are many variations used for obtaining (*b*) and (*c*). In some machines a single device gives both, while in others separate devices are used to obtain each.

(2) THEORY

The performance of a machine can be determined theoretically in many ways, the most widely used being vector and circle diagrams, equivalent circuits, mesh equations and coupled-circuit theory, and tensor analysis.

These methods have been used by different workers in developing the theory of the stator-fed shunt motor. Ridsdale,[12] Schwarz,[15, 27] and others[13, 14, 21] have used vector and circle diagrams in illustrating the operation of the motor. Adkins and Gibbs,[8, 26] and Coulthard[20] used the technique of equivalent circuits in evolving a common theory for all polyphase commutator machines. Klima's paper[29] illustrates the use of mesh equations, while tensors have been successfully used by Kron[4, 9] and Gibbs.[10]

The use of vector and circle diagrams as well as of equivalent circuits, except those developed from exact impedance tensors, usually suffers from approximations. When the machine connections become complex the analysis can best be carried out by tensor methods. The scope of the application of tensor analysis in dealing with the behaviour of commutator machines is very wide indeed, because, by connecting the windings of a machine in different ways, a large number of workable connections can be evolved. Though the basic theory in many of these different workable connections may remain the same, the exact performance equations in different cases may be widely different. When tensors are used, different ways of connecting the windings are taken care of by changing the connection tensor involved, the final equations being thus easily obtained by putting the new connection tensor in place of the old and carrying out the algebraical processes.

In the following, first a simple machine is analysed and its equivalent circuit established before proceeding to the analysis of a general machine.

(2.1) Assumption of Symmetry

In the analyses of the motor it has been assumed that it is symmetrical across both sides of the air-gap. This presumes a sinusoidal current distribution in the commutator winding. Practical experience over a number of years has shown that this assumption, though not theoretically justifiable owing to the action of commutation, does not introduce any appreciable errors.[10]

(2.2) Sign Convention

In subsequent analyses, the positive direction of the induced voltages both in the motor and the regulator secondary windings is taken to be the same as that of the applied voltage (i.e. the transformer effect is neglected). The impressed voltage on the motor brushes is thus considered positive when it opposes the rotor standstill e.m.f. and negative when it is in phase with it. Phase shifts always refer to shifts from the positive direction of the neutral-axis voltage.

The direction of rotation of field in all cases has been arbitrarily taken as clockwise, and for clarity it has been shown by the direction of the arrow in the diagrams concerned.

(2.3) Analysis of a Simple Machine

In an actual machine the phase shift between the rotor impressed voltage and the rotor standstill induced e.m.f. is obtained in a variety of ways. A simple machine connection is shown in Fig. 1. The regulator is supposed to be of a type which can be represented by a transformer with two windings per phase and a variable turns ratio. The phase shift is introduced by a brush shift from the neutral position, by a shift present in the regulator, or by a combination of both. For analytical purposes, as can be seen later, all the phase shift can be considered to be either in the motor or the regulator without affecting the final equations of performance.
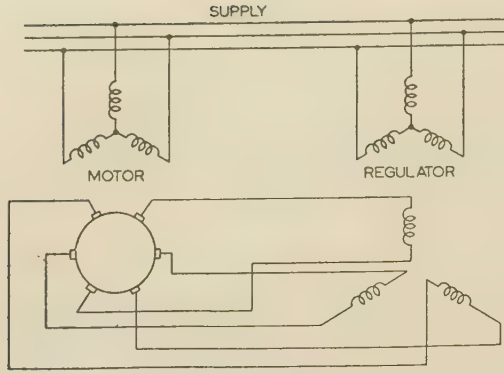
Fig. 1.—A simple machine.

(2.3.1) Analysis when all the Phase Shift is Considered to be in the Motor.

Let the brushes on the motor be shifted by an electrical angle β from the neutral position *in* the direction of rotation, and let the effective turns ratio between the motor windings (secondary/primary) be k_m and between regulator windings (secondary/primary) be k_r, the latter being variable. Fig. 2 shows
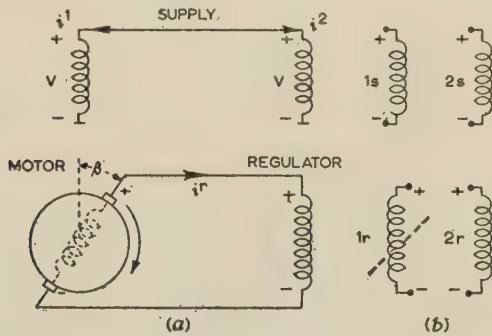
Fig. 2.—Actual and primitive systems of a simple machine.

(a) Actual system.　　　(b) Primitive system.

the actual and primitive systems. In the primitive system the dotted line has been drawn to make the rotation of the winding apparent. The assumption of symmetry (see Section 2.1) permits the consideration of only one phase for analytical purposes. The primitive tensor in terms of actual (not referred) resistances and inductances is given by

$$Z_{mn} = R_{mn} + L_{mn}p + G_{mn}p\theta = R_{mn} + L_{mn}(p - jp\theta)$$

Z_{mn} in the expanded matrix form becomes

$Z_{mn} =$

$m\backslash n$	1s	1r	2s	2r
1s	$R_{s1} + L_{s1}p$	Mp		
1r	$M(p - jp\theta)$	$R_{r1} + L_{r1}(p - jp\theta)$		
2s			$R_{s2} + L_{s2}p$	$M'p$
2r			$M'p$	$R_{r2} + L_{r2}p$

. . . . . (1)

where M and M' are the mutual inductances between the windings 1s and 1r and between 2s and 2r, respectively.

In the torque tensor G_{mn}, m takes the value $1r$ only, since only the winding $1r$ is rotating in normal machine operation. For steady-state operation the operator p equals $j\omega$, and $(p - jp\theta)$ equals $j\omega s$, where s represents the slip of the machine.

If X_{m1} is the reactance of the motor primary winding due to the mutual flux between motor stator and rotor, and X_{m2} is that of the regulator stator winding due to the mutual flux in the regulator,

$$\omega M = \sqrt{(X_{m1}k_m^2 X_{m1})} = k_m X_{m1}; \quad \omega M' = \sqrt{(X_{m2}k_r^2 X_{m2})} = k_r X_{m2}$$

. . . . (2)

Hence

$Z_{mn} =$

$m\backslash n$	1s	1r	2s	2r
1s	$R_{s1} + jX_{s1}$	$jk_m X_{m1}$		
1r	$jk_m sX_{m1}$	$R_{r1} + jsX_{r1}$		
2s			$R_{s2} + jX_{s2}$	$jk_r X_{m2}$
2r			$jk_r X_{m2}$	$R_{r2} + jX_{r2}$

. . . . . (3)

where X_{s1}, X_{r1}, X_{s2}, and X_{r2} are the self-reactances of the windings 1s, 1r, 2s, and 2r, respectively.

The above impedance tensor can be easily recognized as that of an induction motor at slip s with another induction motor at rest.

Let the actual system reference frame components be 1, r, and 2 as indicated in Fig. 2(a). The connection tensor $C^n_{.\nu}$ for transformation from the primitive to the actual system is given by the relation $i^n = C^n_{.\nu}i^\nu$. Hence, by inspection,

$C^n_{.\nu} =$

$n\backslash \nu$	1	r	2
1s	1		
1r		$\varepsilon^{+j\beta}$	
2s			1
2r		-1	

. . . (4)

The impedance tensor for the actual system is given from the equation $Z_{\mu\nu} = C_{.\mu}^{m*}Z_{mn}C^n_{.\nu}$, by

$Z_{\mu\nu} =$

$\mu\backslash\nu$	1	r	2
1	Z_{11}	$k_m Z_{om}\varepsilon^{+j\beta}$	
r	$k_m sZ_{om}\varepsilon^{-j\beta}$	Z_{rr}	$-k_r Z_{or}$
2		$-k_r Z_{or}$	Z_{22}

. . (5)

where

$$Z_{om} = jX_{m1}; \quad Z_{or} = jX_{m2}$$
$$Z_{11} = R_{s1} + jX_{s1} = R_{s1} + jx_{s1} + jX_{m1} = z_{11} + Z_{om}$$
$$Z_{rr} = (R_{r1} + R_{r2}) + j(sX_{r1} + X_{r2})$$
$$= (R_{r1} + R_{r2}) + j(sx_{r1} + x_{r2}) + jk_m^2 sX_{m1} + jk_r^2 X_{m2}$$
$$= z_{rr} + k_m^2 sZ_{om} + k_r^2 Z_{or}$$

. (6)

and

$$Z_{22} = R_{s2} + jX_{s2} = R_{s2} + jx_{s2} + jX_{m2} = z_{22} + Z_{or}$$

and Z refer to self- or mutual-reactances and self- or mutual-impedances respectively, while x and z refer to leakage reactances and leakage impedances, respectively.

The admittance tensor is given by $Y^{\nu\mu} = (Z_{\mu\nu})^{-1}$.

Hence

1	r	2	
$Z_{22}Z_{rr} - k_r^2 Z_{or}^2$	$-k_m Z_{om}Z_{22}\varepsilon^{+j\beta}$	$-k_m k_r Z_{om}Z_{or}\varepsilon^{+j\beta}$	
$-k_m s Z_{om}Z_{22}\varepsilon^{-j\beta}$	$Z_{11}Z_{22}$	$k_r Z_{or}Z_{11}$	$\times \dfrac{1}{D}$
$-k_m k_r s Z_{om}Z_{or}\varepsilon^{-j\beta}$	$k_r Z_{or}Z_{11}$	$Z_{11}Z_{rr} - k_m^2 s Z_{om}^2$	

$$\qquad \cdots \quad (7)$$

where $\quad D = Z_{11}Z_{rr}Z_{22} - Z_{11}k_r^2 Z_{or}^2 - Z_{22}k_m^2 s Z_{om}^2 \quad \cdots \quad (8)$

The actual voltage tensor

μ	
1	V
r	0
2	V

$$v_\mu = \qquad \cdots \qquad (9)$$

Hence from $i^\nu = Y^{\nu\mu}v_\mu$, we have

$$i^1 = \frac{V}{D}(Z_{22}Z_{rr} - k_r^2 Z_{or}^2 - k_m k_r Z_{om}Z_{or}\varepsilon^{+j\beta}) \quad (a)$$

$$i^r = -\frac{V}{D}(k_m s Z_{om}Z_{22}\varepsilon^{-j\beta} - k_r Z_{or}Z_{11}) \quad (b) \qquad (10)$$

$$i^2 = \frac{V}{D}(Z_{11}Z_{rr} - k_m^2 s Z_{om}^2 - k_m k_r s Z_{om}Z_{or}\varepsilon^{-j\beta}) \quad (c)$$

and total supply current I is given by

$$I = i^1 + i^2 \quad \cdots \quad (d)$$

The above relations give the equations of performance of the motor with regulator under steady-state conditions. Under normal operation, k_r and β may vary.

(2.3.2) Analysis when all the Phase Shift is Considered to be in the Regulator.

Keeping all the machine constants the same as in the previous section but considering all the phase shift β to be in the regulator secondary winding, the connection tensor will become

$n\backslash\nu'$	$1'$	r'	$2'$
$1s$	1		
$1r$		1	
$2s$			1
$2r$		$-\varepsilon^{-j\beta}$	

$$C^n_{\nu'} = \qquad \cdots \quad (11)$$

β is considered to be the electrical angle by which the regulator secondary is shifted from its neutral position *against* the direction of rotation of the field.

Using the connection tensor of eqn. (11) and carrying out the regular algebraic processes, the final impedance and admittance tensors become

$\mu'\backslash\nu'$	$1'$	r'	$2'$
$1'$	Z_{11}	$k_m Z_{om}$	
r'	$k_m s Z_{om}$	Z_{rr}	$-k_r Z_{or}\varepsilon^{+j\beta}$
$2'$		$-k_r Z_{or}\varepsilon^{-j\beta}$	Z_{22}

$$Z_{\mu'\nu'} = \qquad \cdots \quad (12)$$

and $\quad Y^{\nu'\mu'} =$

$\nu'\backslash\mu'$	$1'$	r'	$2'$
$1'$	$Z_{22}Z_{rr} - k_r^2 Z_{or}^2$	$-k_m Z_{om}Z_{22}$	$-k_m k_r Z_{om}Z_{or}\varepsilon^{+j\beta}$
r'	$-k_m s Z_{om}Z_{22}$	$Z_{11}Z_{22}$	$k_r Z_{or}Z_{11}\varepsilon^{+j\beta}$
$2'$	$-k_m k_r s Z_{om}Z_{or}\varepsilon^{-j\beta}$	$k_r Z_{or}Z_{11}\varepsilon^{-j\beta}$	$Z_{11}Z_{rr} - k_m^2 s Z_{om}^2$

with $\times \dfrac{1}{D}$

$$\qquad \cdots \quad (13)$$

where D has the same value as in eqn. (8).

The equations of performance are then given by

$$i^{1'} = \frac{V}{D}(Z_{22}Z_{rr} - k_r^2 Z_{or}^2 - k_m k_r Z_{om}Z_{or}\varepsilon^{+j\beta}) \quad (a)$$

$$i^{r'} = -\frac{V}{D}(k_m s Z_{om}Z_{22} - k_r Z_{or}Z_{11}\varepsilon^{+j\beta}) \quad (b) \qquad (14)$$

$$i^{2'} = \frac{V}{D}(Z_{11}Z_{rr} - k_m^2 s Z_{om}^2 - k_m k_r s Z_{om}Z_{or}\varepsilon^{-j\beta}) \quad (c)$$

and the supply current, $I = i^{1'} + i^{2'} \quad \cdots \quad (d)$

(2.3.3) Comparison of the Two Cases.

If the equations of performance as obtained in eqns. (14) are compared with those obtained in eqns. (10), it is evident that the two cases give the same values of motor primary, regulator primary and the total supply currents under conditions of the same speed, same applied voltage, and same shift angle β. However, it is worth remembering that β in Section 2.3.1 was supposed to be the angle of shift of brushes *in* the direction of rotation of flux in the motor (thus giving a lagging motor secondary e.m.f.), while in Section 2.3.2 it was assumed to be a shift in the regulator secondary e.m.f. *against* the direction of rotation of flux in the regulator (thus giving a leading regulator secondary e.m.f.). Thus a lagging phase shift in the brushes is equivalent to a leading phase shift in the regulator by the same amount. There is no exact equivalence in the rotor-circuit currents in the two cases $i^r \neq i^{r'}$, except when β is zero. A simple relationship $i^r = i^{r'}\varepsilon^{-j\beta}$ can be detected, which indicates that the two currents are equal in magnitude but differ in phase by the phase-shift angle β, which is to be expected owing to the nature of the introduction of phase shift in the two cases.

(2.3.4) Ideal No-Load Speeds.

The ideal no-load condition occurs when there is no power output. Thus, in the absence of quadrature e.m.f. in the rotor circuit, under ideal no-load conditions there will be no current flowing in the rotor circuit and the motor will run up to a speed such that the voltage appearing at the brushes will be equal and opposite to the impressed voltage. When quadrature e.m.f. is present, the ideal no-load condition can be supposed to occur when all the no-load secondary current is due to the quadrature

e.m.f. only. Hence if the equation of i^r is expressed in the form of voltage divided by impedance, that component of the numerator which is in phase with the brush standstill open-circuit voltage should be zero under ideal no-load conditions.

From the analysis of Section 2.3.1, i^r on simplification is given by

$$i^r = \frac{- k_m(Z_{om}/Z_{11})V\varepsilon^{-j\beta}(s - m\cos\beta - jm\sin\beta)}{Z_{rr} - k_r^2 Z_{or}^2/Z_{22} - k_m^2 s Z_{om}^2/Z_{11}} \quad . \quad (15)$$

where

$$m = k_r Z_{or} Z_{11}/k_m Z_{om} Z_{22} \quad . \quad . \quad . \quad (16)$$

m has the physical significance of the ratio of the regulator secondary open-circuit voltage to the motor secondary open-circuit voltage, when there is no phase shift between them or when both are referred to their neutral axes. Hence, under ideal no-load condition, since $k_m(Z_{om}/Z_{11})V\varepsilon^{-j\beta}$ is the brush standstill open-circuit voltage,

$$s - m\cos\beta = 0,$$

giving the no-load slip $s = m\cos\beta$, and the no-load speed

$$N_o = N_s(1 - m\cos\beta) \quad . \quad . \quad . \quad (17)$$

where N_s is the synchronous speed of the motor.

In the case of analysis in Section 2.3.2, $i^{r\prime}$, on simplification, becomes

$$i^{r\prime} = \frac{- k_m(Z_{om}/Z_{11})V(s - m\cos\beta - jm\sin\beta)}{Z_{rr} - k_r^2 Z_{or}^2/Z_{22} - k_m^2 s Z_{om}^2/Z_{11}} \quad . \quad (18)$$

where m has the same value as in eqn. (16). In this case, since the brush standstill open-circuit voltage is $k_m(Z_{om}/Z_{11})V$, under ideal no-load condition,

$$s - m\cos\beta = 0$$

giving the same no-load speed as in eqn. (17).

It is obvious that s has, under no-load condition, the same sign as $m\cos\beta$. If m can be considered to be always positive, β then referring to the shift from the positive direction of the secondary e.m.f., s will have the same sign as $\cos\beta$. Hence at sub-synchronous speeds β must be within the limits $\pi/2 > \beta > -\pi/2$ and at super-synchronous speeds within $\pi/2 < \beta < -\pi/2$.

(2.3.5) Power-Factor Correction.

The supply current $I = i^1 + i^2 = i^{1\prime} + i^{2\prime}$ is given from eqn. (10d) or eqn. (14d) on simplification by

$$I = V\left\{\frac{1}{Z_{11}} + \frac{1}{Z_{22}} + \frac{k_m^2(Z_{om}/Z_{11})^2[s - m\cos\beta + m(m - s\cos\beta)]}{Z_{rr} - k_r^2 Z_{or}^2/Z_{22} - k_m^2 s Z_{om}^2/Z_{11}}\right.$$
$$\left. + j\frac{k_m^2(Z_{om}/Z_{11})^2 m(s - 1)\sin\beta}{Z_{rr} - k_r^2 Z_{or}^2/Z_{22} - k_m^2 s Z_{om}^2/Z_{11}}\right\} \quad . \quad (19)$$

It is obvious that the last term on the right-hand side of eqn. (19) is the phase compensation term as it contains the quadrature e.m.f. component. Since, in general, the first three terms give lagging current, if a power-factor improvement is desired, $m(s - 1)\sin\beta$ must always be positive. $(s - 1)$ is negative over the entire speed range except at $s = 1$, when, of course, the quadrature component has no effect on the supply current. Hence $m\sin\beta$ must be negative over the entire speed range to give phase compensation, and as m is always positive in the convention adopted in the paper, β must always lie within the limits $\pi < \beta < 2\pi$. Comparing this with the limits for sub- and super-synchronous operations established in the previous Section, a power-factor correction at sub-synchronous speeds requires β to be in the fourth quadrant and at super-synchronous speeds in the third quadrant.

Stated more simply, while $m\cos\beta$ must change sign from positive to negative under the two conditions of operation $m\sin\beta$ must always be negative. This means that, though the speed-regulation component of the regulating voltage should change sign with respect to the rotor standstill e.m.f. in order to permit both sub- and super-synchronous operations, the quadrature component should maintain the same sign to give compensation in both the ranges.

It is thus clear that a fixed shift of the brushes, coupled with a regulator which can only give voltages positive and negative to the rotor standstill neutral-axis e.m.f., can give correction only in the sub-synchronous or the super-synchronous speed range but not in both. A small shift (less than 90°) of the brushes against the direction of rotation (β in fourth quadrant) will give a power-factor improvement in the sub-synchronous range, while a similar shift in the direction of rotation will improve the power factor at super-synchronous speeds (since the total shift β is the sum of the shift due to brush position and a shift of $+180°$, owing to reversal of regulator polarity and is therefore in third quadrant).

The simplest way of obtaining a quadrature component of voltage which does not change sign in the sub- and super-synchronous speed ranges is from an auxiliary apparatus, e.g. a compensating transformer, or an auxiliary winding on the stator. A biased double regulator also gives such a voltage.[34] Since $m\sin\beta$ should be negative, the quadrature e.m.f. obtained in the regulator should lag the positive direction of the neutral-axis voltage, because when the shift is in the regulator (Section 2.3.2) β is the angle of lead from the neutral position. The quadrature e.m.f. injected is limited by consideration of the permissible no-load current in the rotor circuit, since, as expressed earlier, under ideal no-load condition the rotor-circuit current is entirely due to the quadrature e.m.f.

(2.4) Equivalent Circuit of the Simple Machine

Though the performance of a machine can be calculated from the equations of performance, it is sometimes useful to establish equivalent circuits to facilitate numerical solutions. Before deriving exact equivalent circuits, it will be shown how the existing circuit can be derived from the tensor equations.

(2.4.1) Coulthard's Circuit.

One of the equivalent circuits in use at present is that proposed by Coulthard,[20] and also by Taylor,[11] which gives only the current taken through the motor. This circuit can easily be derived from the equation of motor primary current [eqn. (10a) or (14a)]. We have, on simplification,

$$i^1 = i^{1\prime} = \frac{Z_{rr} - k_r^2 Z_{or}^2/Z_{22} - k_m k_r Z_{om}(Z_{or}/Z_{22})\varepsilon^{+j\beta}}{Z_{11}Z_{rr} - k_r^2(Z_{or}^2/Z_{22})Z_{11} - k_m^2 s Z_{om}^2}V. \quad (20)$$

which, from eqns. (6), reduces to

$$i^1 = (V - i^1 z_{11})\left[\frac{1}{Z_{om}} + \frac{k_m^2(s - b\varepsilon^{+j\beta})}{z_{rr} + k_r^2(Z_{or}/Z_{22})z_{22}}\right] \quad . \quad (21)$$

where b is the ratio of the regulator neutral-axis open-circuit voltage (v_k) to the motor secondary standstill neutral-axis e.m.f. v_2,

i.e.

$$b = \frac{k_r(Z_{or}/Z_{22})V}{k_m(V - i_1 z_{11})} \quad . \quad . \quad . \quad . \quad (22)$$

But $z_{rr} + k_r^2(Z_{or}/Z_{22})z_{22}$ is the sum of the leakage impedance of the rotor circuit (i.e. motor and regulator secondary windings) and the referred leakage impedance of the regulator primary

$r_t + jx_t$ represent the total effective leakage impedance of the regulator referred to its secondary side,

$$z_{rr} + k_r^2(Z_{or}/Z_{22})z_{22} = (R_{r1} + r_t) + j(sx_{r1} + x_t) \quad . \quad (23)$$

From eqns. (21) and (23) the equivalent circuit becomes as shown in Fig. 3. In this circuit, from eqn. (22) $b = v_k/v_2$.
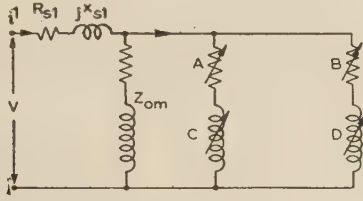
Fig. 3.—Existing equivalent circuit for the simple machine.

Resistor A: $\dfrac{R_{r1} + r_t}{k_m^2(s - b \cos \beta)}$

Resistor B: $-\dfrac{sx_{r1} + x_t}{k_m^2 b \sin \beta}$

Inductor C: $j\dfrac{sx_{r1} + x_t}{k_m^2(s - b \cos \beta)}$

Inductor D: $j\dfrac{R_{r1} + r_t}{k_m^2 b \sin \beta}$

From the convention in Section 2.2, v_k is positive when it opposes v_2. Sometimes the positive direction of v_k is taken as that which aids v_2 in a series circuit. Under such a convention, b has to be replaced by $-b$ in the equivalent circuit of Fig. 3, which then takes the same form as the Coulthard's[20] or Taylor's[11] circuit.

(2.4.2) Limitations of Coulthard's Circuit.

Apart from giving only the motor current and thus requiring additional calculations for finding the total supply current, a serious disadvantage of this circuit is the dependence of b on the current i^1 [eqn. (22)]. In many cases, v_2 is approximated by the open-circuit standstill motor secondary voltage, i.e. $v_2 = Vk_m(Z_{om}/Z_{11})$, giving $b = k_r Z_{or} Z_{11}/k_m Z_{om} Z_{22}$. However, this approximation is not justifiable, because the final equation for i^1 using this approximate value of b becomes

$$i^1 =$$

$$\frac{\left[z_{rr} - k_r^2 Z_{or}^2/Z_{22} - k_m k_r Z_{om}(Z_{or}/Z_{22})\varepsilon^{+j\beta} - k_m k_r(Z_{or}/Z_{22})z_{11}\varepsilon^{+j\beta}\right]}{\left[Z_{11}Z_{rr} - k_r^2(Z_{or}^2 Z_{22})Z_{11} - k_m^2 s Z_{om}^2 - k_m k_r(Z_{or}/Z_{22})Z_{11}z_{11}\varepsilon^{+j\beta}\right]}$$
$$\times V \quad . \quad . \quad . \quad (24)$$

If eqn. (24) is compared with eqn. (20), it is seen that the last terms in the numerator as well as in the denominator for the expression of i^1 in eqn. (24) are superfluous. Thus for exact representation of the machine performance by the equivalent circuit of Fig. 3, b must be considered as a function of the motor primary current. This makes any numerical solution very difficult, and for exact calculation it is necessary to revert to eqn. (20).

(2.4.3) Exact Equivalent Circuit of the Coulthard Type.

However, an exact practical circuit of the same type, i.e. one which gives the current through the motor only but does not suffer from the limitation outlined in the last Section (i.e. dependence of b on i^1) can be easily established. On simplification of eqn. (10a) or (14a), the general equation for the motor primary current can be written

$$i^1 = V\left[\frac{1}{Z_{11}} + \frac{k_m^2(Z_{om}/Z_{11})^2\left(s - \dfrac{k_r Z_{or} Z_{11}}{k_m Z_{om} Z_{22}}\varepsilon^{+j\beta}\right)}{z_{rr} + k_m^2 s(Z_{om}/Z_{11})z_{11} + k_r^2(Z_{or}/Z_{22})z_{22}}\right] \quad . \quad (25)$$

$z_{rr} + k_m^2 s(Z_{om}/Z_{11})z_{11} + k_r^2(Z_{or}/Z_{22})z_{22}$ is the sum of the leakage impedances of the motor and regulator referred to their secondary sides. Hence, if $(r_m + jsx_m)$ is the total leakage impedance of the motor referred to its secondary side, and $(r_t + jx_t)$ is the total leakage impedance of the regulator referred to its secondary side,

$$(r_m + jsx_m) + (r_t + jx_t)$$
$$= z_{rr} + k_m^2 s(Z_{om}/Z_{11})z_{11} + k_r^2(Z_{or}/Z_{11})z_{22} \quad . \quad (26)$$

Let $$k_m' = k_m(Z_{om}/Z_{11}) \quad . \quad . \quad . \quad . \quad (27)$$

k_m' having the physical significance of the ratio of motor secondary open-circuit voltage (neutral axis) to the applied primary voltage.

Then eqn. (25) reduces to

$$i^1 = V\left[\frac{1}{Z_{11}} + \frac{k_m'^2(s - m \cos \beta)}{(r_m + r_t) + j(sx_m + x_t)} - \frac{jk_m'^2 m \sin \beta}{(r_m + r_t) + j(sx_m + x_t)}\right]$$
$$. \quad . \quad . \quad . \quad (28)$$

where m has the same value and significance as in eqn. (16). The equivalent circuit of Fig. 4 is a direct representation of
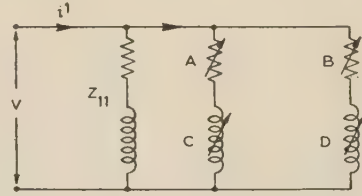
Fig. 4.—Exact equivalent circuit of the Coulthard type.

Resistor A: $\dfrac{r_m + r_t}{k_m'^2(s - m \cos \beta)}$

Resistor B: $-\dfrac{sx_m + x_t}{k_m'^2 m \sin \beta}$

Inductor C: $j\dfrac{sx_m + x_t}{k_m'^2(s - m \cos \beta)}$

Inductor D: $j\dfrac{r_m + r_t}{k_m'^2 m \sin \beta}$

Impedance Z_{11}: $R_{s1} + jx_{s1} + Z_{om}$

eqn. (28). This circuit does not suffer from the dependence of b on i^1 because the corresponding quantity m in this circuit, as shown in eqn. (16), is independent of the motor current. Moreover, the impedance branches are all in parallel, which facilitates numerical solutions.

(2.4.4) Equivalent Circuit giving both Motor and Regulator Currents

The equivalent circuits given by Figs. 3 and 4 suffer from the common defect of giving only the motor current. However, a circuit of the type of Fig. 4 can be easily established to give both the motor and regulator currents. From eqn. (10c) or (14c), after simplification, the regulator primary current is given by

$$i^2 = V\left[\frac{1}{Z_{22}} + \frac{k_r' k_m'(m - s \cos \beta)}{(r_m + r_t) + j(sx_m + x_t)} + j\frac{k_r' k_m' s \sin \beta}{(r_m + r_t) + j(sx_m + x_t)}\right]$$
$$. \quad . \quad . \quad (29)$$

where $$k_r' = k_r Z_{or}/Z_{22} \quad . \quad . \quad . \quad . \quad (30)$$

k_r' has the physical significance of the ratio of the open-circuit regulator secondary voltage (neutral axis) to the applied primary voltage. The equivalent circuit to represent eqn. (29) is shown in Fig. 5. The complete equivalent circuit is obtained by combining the two circuits of Figs. 4 and 5 and is shown in Fig. 6.
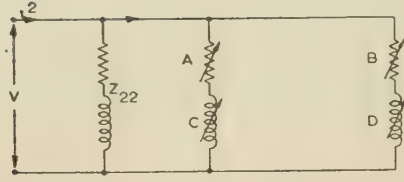
Fig. 5.—Equivalent circuit for the current taken through the regulator.

Resistor A: $\dfrac{r_m + r_t}{k_t' k_u'(m - s\cos\beta)}$

Resistor B: $\dfrac{sx_m + x_t}{k_t' k_m' s\sin\beta}$

Inductor C: $j\dfrac{sx_m + x_t}{k_t' k_m'(m - s\cos\beta)}$

Inductor D: $-j\dfrac{r_m + r_t}{k_t' k_m' s\sin\beta}$
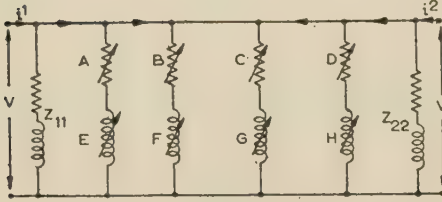
Fig. 6.—Equivalent circuit giving both motor and regulator currents.

Resistor A: $\dfrac{r_m + r_t}{k_m'^2(s - m\cos\beta)}$ Inductor E: $j\dfrac{sx_m + x_t}{k_m'^2(s - m\cos\beta)}$

Resistor B: $-\dfrac{sx_m + x_t}{k_m'^2 m\sin\beta}$ Inductor F: $j\dfrac{r_m + r_t}{k_m'^2 m\sin\beta}$

Resistor C: $\dfrac{r_m + r_t}{k_t' k_m'(m - s\cos\beta)}$ Inductor G: $j\dfrac{sx_m + x_t}{k_t' k_m'(m - s\cos\beta)}$

Resistor D: $\dfrac{sx_m + x_t}{k_t' k_m' s\sin\beta}$ Inductor H: $-j\dfrac{r_m + r_t}{k_t' k_m' s\sin\beta}$

This circuit gives both the motor primary and the regulator primary currents and hence the total supply current. If only the supply current is needed, and knowledge of currents taken separately through the motor and the regulator is not required, the circuit of Fig. 6 can be reduced to the form shown in Fig. 7.
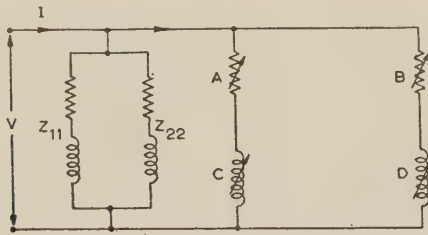
Fig. 7.—Equivalent circuit giving the total supply current.

Resistor A: $\dfrac{r_m + r_t}{k_m'^2[(s - m\cos\beta) + m(m - s\cos\beta)]}$

Resistor B: $\dfrac{sx_m + x_t}{k_m'^2 m(s - 1)\sin\beta}$

Inductor C: $j\dfrac{sx_m + x_t}{k_m'^2[(s - m\cos\beta) + m(m - s\cos\beta)]}$

Inductor D: $-j\dfrac{r_m + r_t}{k_m'^2 m(s - 1)\sin\beta}$

It is interesting to note that the expression

$$k_m'^2[(s - m\cos\beta) + m(m - s\cos\beta) + jm(s - 1)\sin\beta]$$

is, as expected, the exact transformation ratio between the rotor and supply impedances. If the supply voltage is V, the voltage

in the rotor circuit when the phase shift is considered to be i the brushes is $(k_m' s\varepsilon^{-j\beta} - k_r')V$. Hence the voltage transforma tion ratio (supply : rotor) is

$$1 : (k_m' s\varepsilon^{-j\beta} - k_r')$$

Therefore, the current transformation ratio (supply : rotor) i

$$(k_m'\varepsilon^{+j\beta} - k_r') : 1$$

and hence the impedance transformation ratio (supply : rotor

is $$1 : (k_m' s\varepsilon^{-j\beta} - k_r')(k_m'\varepsilon^{+j\beta} - k_r')$$

which, on simplification, yields

$$1 : k_m'^2[(s - m\cos\beta) + m(m - s\cos\beta) + jm(s - 1)\sin\beta]$$

(2.4.5) Equivalent Circuits from Impedance Tensor.

In the preceding Sections, equivalent circuits have beer derived from actual performance equations. In systems where the performance equations are more complex, such derivation not only takes time but becomes very difficult to establish It is, however, possible to establish an equivalent circuit from the final impedance tensor of the system.[4,9,10] Wher the impedance tensor is asymmetrical as in eqns. (5) and (12) it is necessary either to make it symmetrical by easy transforma tions[4,10] or to use the principle of the ideal transfer network with asymmetrical voltage and current transformation ratios A number of equivalent circuits for the simple machine under consideration can be easily established using these techniques.[3 Two such circuits, using ideal transfer networks, are shown ir Figs. 8 and 9. As can be seen, they refer to the impedance tensor in eqn. (5). Similar circuits can also be established fo the machine with the impedance tensor given by eqn. (12). The circuit of Fig. 8 has been named the 'orthodox circuit' because it is a combination of an orthodox induction-motor circuit with that of a transformer. The circuit of Fig. 9, however, is termed the 'exact practical circuit' owing to its similarity in principle to Morris's[30] exact practical circuit for a transformer.

(2.5) General Analysis

The auxiliary winding on the stator for power-factor improve ment and the double regulator for speed regulation introduce complications which make the simple analysis of Section 2.3.1 o 2.3.2 no longer exactly applicable. Since the auxiliary winding i in the motor main field, it introduces mutual coupling with both the stator main winding and the rotor winding. Owing to it internal winding connections, a double regulator cannot be repre sented by an impedance tensor of the transformer type, i.e. by one representing two mutually coupled windings. To make th analysis more general, consideration is given to the case of motor which has its brushes shifted from its neutral position, a auxiliary stator winding connected in its rotor circuit, and biased double regulator for speed regulation and further powe factor control.

Let θ be the brush shift against the direction of rotation, an let the auxiliary-winding e.m.f. lag the stator e.m.f. by γ (neg lecting transformer effect). The biased double regulator i considered for the sake of generality to be composed of tw completely dissimilar single regulators.

The primitive and actual systems for the general machine ar shown in Fig. 10. It may be noted that in the primitive system the biased double regulator is left with its internal termina connected, and is thus treated as a four-terminal network wit currents i^{2r} and i^{2s} flowing through the output and inpu terminal, respectively.
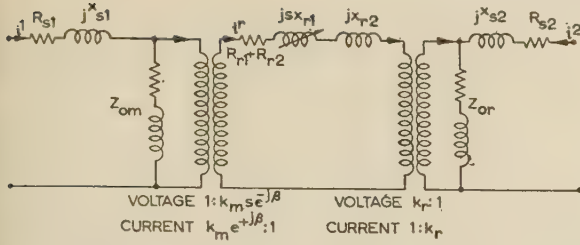
Fig. 8.—Orthodox circuit for the simple machine using the ideal transfer network.

Fig. 9.—Exact practical equivalent circuit for the simple machine.

$$R = R_{r1} + R_{r2} + k_r'^2 R_{s2}; \quad R' = k_m'^2 R_{s1}$$
$$x = x_{r2} + k_r'^2 x_{s2}; \quad x' = x_{r1} + k_m'^2 x_{s1}$$

It has been shown in an earlier paper[34] that the impedance tensor for a biased double regulator with dissimilar component elements treated as a four-terminal network is given by

$$Z_{\alpha\beta} = \begin{array}{c|c|c} \alpha \backslash \beta & 2s & 2r \\ \hline 2s & \dfrac{Z_{a1}Z_{b1}}{Z_{a1}+Z_{b1}} & \dfrac{Z_{a1}Z_{b1}}{Z_{a1}+Z_{b1}}\left(k_a\dfrac{Z_{oa}}{Z_{a1}}\varepsilon^{+j\alpha} + k_b\dfrac{Z_{ob}}{Z_{b1}}\varepsilon^{-j\alpha}\right) \\ \hline 2r & \dfrac{Z_{a1}Z_{b1}}{Z_{a1}+Z_{b1}}\left(k_a\dfrac{Z_{oa}}{Z_{a1}}\varepsilon^{-j\alpha} + k_b\dfrac{Z_{ob}}{Z_{b1}}\varepsilon^{+j\alpha}\right) & Z_{a2}+Z_{b2} - \dfrac{k_a^2 Z_{oa}^2 + k_b^2 Z_{ob}^2 - 2k_a k_b Z_{oa} Z_{ob}\cos 2\alpha}{Z_{a1}+Z_{b1}} \end{array} \quad . \quad (31)$$

where α = Electrical angle by which the regulator shaft is rotated from its neutral position.

k_a and k_b = Effective turns ratios of the two component single regulators.

Z_{a1}, Z_{a2}, Z_{b1}, and Z_{b2} = Self-impedances of the four windings of the regulator ($1a$ and $1b$ represent the two primary windings, and $2a$ and $2b$ the two secondary windings).

Z_{oa} and Z_{ob} = Impedances of the windings $1a$ and $1b$ due to the mutual fluxes between $1a$ and $2a$, and between $1b$ and $2b$, respectively.

If $Z_p = Z_{a1}Z_{b1}/(Z_{a1}+Z_{b1})$ = Self-impedance of the two primary windings in parallel.

$Z_R = Z_{a2} + Z_{b2}$ = Self-impedance of the two secondary windings in series.

$$k_a' = k_a Z_{oa}/Z_{a1}, \text{ and } k_b' = k_b Z_{ob}/Z_{b1} \quad . \quad . \quad . \quad . \quad (32)$$

Eqn. (31) reduces to

Fig. 10.—Actual and primitive systems for the general machine.
(a) Actual system.   (b) Primitive system.

$$Z_{\alpha\beta} = \begin{array}{c|c|c} \alpha \backslash \beta & 2s & 2r \\ \hline 2s & Z_p & Z_p(k_a'\varepsilon^{+j\alpha} + k_b'\varepsilon^{-j\alpha}) \\ \hline 2r & Z_p(k_a'\varepsilon^{-j\alpha} + k_b'\varepsilon^{+j\alpha}) & Z_R - Z_p\left(k_a'^2\dfrac{Z_{a1}}{Z_{b1}} + k_b'^2\dfrac{Z_{b1}}{Z_{a1}} - 2k_a'k_b'\cos 2\alpha\right) \end{array} \quad . \quad . \quad . \quad (33)$$

The primitive impedance tensor for the general machine is found by inspection:

$$Z_{mn} = \begin{array}{c|c|c|c|c|c} m \backslash n & 1s & 1r & 2s & 2r & 3s \\ \hline 1s & R_{s1}+jX_{s1} & j\omega M_1 & & & j\omega M_3 \\ \hline 1r & j\omega s M_1 & R_{r1}+js X_{r1} & & & j\omega s M_4 \\ \hline 2s & & & Z_p & Z_p(k_a'\varepsilon^{+j\alpha} + k_b'\varepsilon^{-j\alpha}) & \\ \hline 2r & & & Z_p(k_a'\varepsilon^{-j\alpha} + k_b'\varepsilon^{+j\alpha}) & Z_R - Z_p\left(k_a'^2\dfrac{Z_{a1}}{Z_{b1}} + k_b'^2\dfrac{Z_{b1}}{Z_{a1}} - 2k_a'k_b'\cos 2\alpha\right) & \\ \hline 3s & j\omega M_3 & j\omega M_4 & & & R_{s3}+jX_{s3} \end{array} \quad (34)$$

The only winding rotating in the system is the motor secondary winding $1r$, and hence terms containing s appear in the components Z_{1r1s}, Z_{1r1r}, and Z_{1r3s}, but nowhere else.

If X_{m1} is the reactance of the motor primary winding due to the mutual flux between the stator and the rotor, and k_m and k_x are the effective turns ratios between the motor secondary and primary windings and between the motor auxiliary and primary windings, respectively,

$$\omega M_1 = k_m X_{m1}, \quad \omega M_3 = k_x X_{m1}, \quad \text{and} \quad \omega M_4 = k_m k_x X_{m1} \ . \quad (35)$$

The primitive impedance tensor then becomes

$$Z_{mn} =$$

$m \backslash n$	$1s$	$1r$	$2s$	$2r$	$3s$
$1s$	$R_{s1}+jX_{s1}$	$jk_m X_{m1}$			$jk_x X_{m1}$
$1r$	$jk_m s X_{m1}$	$R_{r1}+jsX_{r1}$			$jsk_m k_x X_{m1}$
$2s$			Z_p	$Z_p(k'_a \varepsilon^{+j\alpha} + k'_b \varepsilon^{-j\alpha})$	
$2r$		$Z_p(k'_a \varepsilon^{-j\alpha} + k'_b \varepsilon^{+j\alpha})$	$Z_R - Z_p\left(k'^2_a \dfrac{Z_{a1}}{Z_{b1}} + k'^2_b \dfrac{Z_{b1}}{Z_{a1}} - 2k'_a k'_b \cos 2\alpha\right)$		
$3s$	$jk_x X_{m1}$	$jk_m k_x X_{m1}$			$R_{s3}+jX_{s3}$

$$(36)$$

The connection tensor is established by comparison of the actual system currents with the primitive system currents, and is given by inspection as

$$C^n_{.\nu} =$$

$n \backslash \nu$	1	r	2
$1s$	1		
$1r$		$\varepsilon^{-j\theta}$	
$2s$			1
$2r$		-1	
$3s$		$-\varepsilon^{+j\gamma}$	

$$. \ . \ . \ (37)$$

Since the actual system impedance tensor $Z_{\mu\nu}$ is $C_\mu^{.m*} Z_{mn} C^n_{.\nu}$,

$$Z_{\mu\nu} =$$

$\mu \backslash \nu$	1	r	2
1	Z_{11}	Z_{1r}	
r	Z_{r1}	Z_{rr}	$-Z_{r2}$
2		$-Z_{2r}$	Z_{22}

$$. \ . \ . \ (38)$$

where
$$
\begin{aligned}
Z_{11} &= R_{s1}+jX_{s1} = R_{s1}+jx_{s1}+jX_{m1} \\
&= z_{11}+jX_{m1} \\
Z_{1r} &= jX_{m1}(k_m \varepsilon^{-j\theta} - k_x \varepsilon^{+j\gamma}) \\
Z_{r1} &= jX_{m1}(k_m s \varepsilon^{+j\theta} - k_x \varepsilon^{-j\gamma}) \\
Z_{rr} &= R_{r1}+jsX_{r1}+R_{s3}+jX_{s3}+Z_R-Z_p \\
&\quad \left(k'^2_a \frac{Z_{a1}}{Z_{b1}} + k'^2_b \frac{Z_{b1}}{Z_{a1}} - 2k'_a k'_b \cos 2\alpha\right) \\
&\quad - jk_m k_x X_{m1}[\varepsilon^{-j(\gamma+\theta)} + s\varepsilon^{j(\gamma+\theta)}] \\
Z_{r2} &= Z_p(k'_a \varepsilon^{-j\alpha} + k'_b \varepsilon^{+j\alpha}) \\
Z_{2r} &= Z_p(k'_a \varepsilon^{+j\alpha} + k'_b \varepsilon^{-j\alpha}) \\
\text{and} \quad Z_{22} &= Z_p
\end{aligned}
\quad (39)
$$

from which the admittance tensor becomes

$$Y^{\nu\mu} =$$

$\nu \backslash \mu$	1	r	2
1	$Z_{rr}Z_{22}-Z_{r2}Z_{2r}$	$-Z_{1r}Z_{22}$	$-Z_{1r}Z_{r2}$
r	$-Z_{r1}Z_{22}$	$Z_{11}Z_{22}$	$Z_{11}Z_{r2}$
2	$-Z_{r1}Z_{2r}$	$Z_{11}Z_{2r}$	$Z_{11}Z_{rr}-Z_{r1}Z_{1r}$

$$\times \frac{1}{D} \qquad\qquad . \ . \ . \ . \ (40)$$

where
$$D_1 = Z_{11}Z_{rr}Z_{22} - Z_{11}Z_{r2}Z_{2r} - Z_{1r}Z_{r1}Z_{22} \ . \quad (41)$$

The voltage tensor for the actual system is

μ	
$v_\mu = \ \ 1$	V
r	0
2	V

Hence,
$$
\begin{aligned}
i^1 &= \frac{V}{D_1}(Z_{rr}Z_{22} - Z_{r2}Z_{2r} - Z_{1r}Z_{r2}) \\
i^r &= \frac{V}{D_1}(Z_{11}Z_{r2} - Z_{r1}Z_{22}) \\
i^2 &= \frac{V}{D_1}(Z_{11}Z_{rr} - Z_{r1}Z_{1r} - Z_{r1}Z_{2r})
\end{aligned}
\right\} \quad . \ \therefore \ (42)
$$

The actual values of i^ν can be calculated for any given values of θ, γ, k_m, k_x, k_a, k_b, α and s. In general, the only variables in any particular machine are α (depending on the regulator shaft position) and s (depending on load conditions), while θ, γ, k_a, k_b, k_x, and k_m are all fixed. Thus the performance equations can be expressed in terms of α and s if numerical values of all other parameters, supposed constant, are substituted in eqns. (42). From the general impedance tensor of eqn. (38), the actual impedance tensor to suit any particular machine arrangement can be easily derived. The absence of an auxiliary winding can be taken into account by putting $k_x = 0$, and $R_{s3} + jX_{s3} = 0$, and the absence of any brush shift by putting $\theta = 0$. The use of a biased double regulator with similar primary, or secondary windings, or the use of an unbiased double regulator, can be taken into account by making suitable changes in the regulator parameters, i.e. by putting $Z_{a1} = Z_{b1}$ (similar primary windings), or $Z_{a2} = Z_{b2}$ (similar secondary windings), or $Z_{a1} = Z_{b1}$, $Z_{a2} = Z_{b2}$, and $k_a = k_b$ (unbiased regulator). A two-winding regulator can be covered by making the self-

impedance of one of the primary windings infinite and that of one of the secondary windings zero.

Iron losses in the machine have so far been neglected. If experimentally determined, they can be added separately to the machine power input as determined by the theory. A better method, universally adopted in the theory of iron-cored transformers and induction machines, is to consider the mutual-impedance terms to have power components in addition to mutual-inductance components. Thus, in the case of the motor, jX_{m1} is replaced by Z_{om}, where Z_{om} includes a resistive component. Similarly Z_{oa} and Z_{ob}, the mutual-impedance terms in the regulator, can be considered to include iron losses. As they are sums of leakage and mutual-impedance components, the self-impedance terms in the theory thus include resistive components to account for iron losses in addition to the actual winding resistances.

(2.6) Equivalent Circuit for the General Machine

The value of equivalent circuits for machine performance calculations cannot be over-emphasized. The use of tensors in establishing equivalent circuits for complicated machine connections has been widespread,[4, 9, 10] and they scarcely need elaboration. The impedance tensor for the general machine [eqn. (38)] is asymmetrical, and the asymmetry is such as to make the formulation of an equivalent symmetrical circuit by easy transformations very difficult. However, when the concept of an ideal transfer network with different voltage and current transformation ratios is used, no equivalent symmetrical circuit is necessary. Using this concept both the orthodox and the exact practical types of equivalent circuits can be easily established.[33] However, as the practical circuit is always easier to manipulate, for both the experimental measurement of parameters and the calculation of performance, only this type will be discussed. The circuit for the general machine is shown in Fig. 11, where

$$k_1 = (k_m s \varepsilon^{+j\theta} - k_x \varepsilon^{-j\gamma}) Z_{om}/Z_{11};$$

$$k_2 = (k_m \varepsilon^{-j\theta} - k_x \varepsilon^{+j\gamma}) Z_{om}/Z_{11};$$

$$k_3 = k_a' \varepsilon^{-j\alpha} + k_b' \varepsilon^{+j\alpha}; \quad k_3^* = k_a' \varepsilon^{+j\alpha} + k_b' \varepsilon^{-j\alpha}$$

and $Z' = Z_{rr} - k_1 k_2 Z_{11} - k_3 k_3^* Z_p$, where Z_{rr} is the rrth component of the general impedance tensor [eqn. (38)].

Z' on simplification becomes

$$(R_{r1} + jsx_{r1}) + (R_{s3} + jsx_{s3})$$
$$+ \left\{ k_m^2 s + k_x^2 - k_m k_x [\varepsilon^{-j(\gamma+\theta)} + s\varepsilon^{j(\gamma+\theta)}] \right\}$$
$$\frac{Z_{om}}{Z_{11}} z_{11} + z_R + \left(k_a^2 \frac{Z_{oa}}{Z_{a1}} z_{a1} + k_b^2 \frac{Z_{ob}}{Z_{b1}} z_{b1} \right) \quad (43)$$

= Leakage impedance of (motor rotor + motor auxiliary + referred value of motor primary + regulator secondaries + referred value of regulator primary).

= Total leakage impedance of the machine referred to the rotor circuit.

(2.6.1) Measurement of Parameters.

The parameters of the equivalent circuit of Fig. 11 are as follows:

Z_{11} = Self-impedance of the motor primary winding.
Z_p = Self-impedance of the regulator primary windings in parallel.
k_1 = Voltage ratio of the ideal transfer network representing the motor.

k_2 = Current ratio of the ideal transfer network representing the motor.
k_3 = Voltage ratio of the ideal transformer representing the regulator.
k_3^* = Current ratio of the ideal transformer representing the regulator.
Z' = Total leakage impedance of the motor and regulator referred to the rotor circuit.
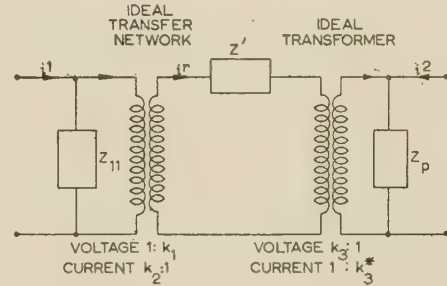
Fig. 11.—Equivalent circuit for the general machine.

In normal induction machines, an open-circuit and a short-circuit test are sufficient for the measurement of equivalent circuit parameters. It will be shown that the same is true for this machine.

Open-circuit test.—In this test the rotor circuit is open-circuited and the motor and regulator primary windings are connected to the normal-voltage normal-frequency supply. The rotor must be locked to prevent rotation due to currents in the coils short-circuited by the brushes. If, in the test,

V = Supply voltage, volts per phase.
i^1 = Motor primary current, amperes per phase.
i^2 = Regulator primary current, amperes per phase.
V_{bx} = Voltage across the motor brushes and auxiliary winding in series, volts per phase.
and V_r = Voltage across the regulator secondary terminals, volts per phase.

From the equivalent circuit,

$$Z_{11} = V/i^1 \text{ ohms}; \quad Z_p = V/i^2 \text{ ohms}; \quad k_1 = V_{bx}/V; \text{ and } k_3 = V_r/V$$
$$\cdots \cdots \cdots (44)$$

Short-circuit test.—In this test, the rotor circuit of the machine is broken at any convenient point and the terminals are connected to a balanced 3-phase normal-frequency supply, while the motor and regulator primary windings are short-circuited. If, in the test,

V_{sc} = Voltage in the rotor circuit, volts per phase.
i_{sc}^1 = Short-circuit current in the motor primary winding, amperes per phase.
i_{sc}^2 = Short-circuit current in the double regulator primary winding, amperes per phase.
i_{sc}^r = Current in the rotor circuit, amperes per phase.

then $Z' = V_{sc}/i_{sc}^1 \text{ ohms}; \quad k_2 = i_{sc}^1/i_{sc}^r; \text{ and } k_3^* = i_{sc}^2/i_{sc}^r$ (45)

Since all the parameters are complex, the phases of all the voltages and currents with respect to the supply voltages in the tests must be measured. The parameters Z', due to the presence of the slip-dependent portion of leakage impedance, and k_1, since it contains a slip-dependent term, are functions of speed; and to find their dependence on speed, the values should be measured at a number of speeds. However, as the dependence is linear,

measurement at two different slips is sufficient. When taking measurements at any speed other than when $s = 1$ (standstill) the motor rotor should be driven by an auxiliary drive. In practice, the measurement can be slightly modified by separating the motor and regulator parameters. The latter can be measured with great accuracy, as the regulator is a static piece of equipment.[34] The motor parameters can be measured by tests similar in principle to the open-circuit and short-circuit tests described above.

(2.7) Limitations of Theory and Suggested Modifications

The theory and equivalent circuits as developed in earlier Sections have inherent limitations, since, in their development, no consideration has been given to the following:

(a) Effect of non-linearity in the rotor circuit due to carbon-brush contact resistance.
(b) Effect of short-circuited coils under the brushes.
(c) Effect of the slip-independent portion of motor-armature leakage reactance.
(d) Effect of variation of no-load losses with speed.
(e) Effect of unbalance, harmonics and non-sinusoidal flux densities.

(2.7.1) Non-Linearity in Rotor Circuit due to Carbon-Brush Contact Resistance.

In the shunt motor, the brush contact resistance constitutes the major part of the rotor-circuit resistance. The carbon-brush contact resistance is known[3, 6] to have non-linear properties and is a function of the current flowing through the contact. The effect of this non-linearity has been studied by Gibbs[24] for the Schrage motor. The only way of taking this effect in consideration in the general theory and the equivalent circuit is to resolve the motor secondary resistance R_{r1} into two components—R'_{r1} representing the winding resistance per phase and R_b representing the brush contact resistance per phase. Since R_b is a function of i^r, at any given slip and regulator setting,

$$i^r = \frac{\text{A constant}}{\text{Function } (i^r) + \text{Another constant}}$$

It is obvious that the solution for i^r becomes involved. However, successive approximation can be used to good purpose in obtaining a fairly accurate value of i^r, if the variation of R_b with i^r is known.

In addition to non-linear resistance, the carbon contact has other properties which may affect the machine behaviour. It has a negative temperature coefficient of resistance, and at high peripheral speeds (above about $4\,000\,\text{ft/min}$)[6] shows a dependence on the speed of the contact surface. Many of the divergences in theoretical performance calculations and experimental results in machines where carbon contact plays an important part can be traced to the susceptibilities of the contact to various outside and inside influences (e.g. temperature, humidity, current density, peripheral speed, etc.).

(2.7.2) The Short-Circuited Coils under the Brushes.

In commutator machines a number of coils are always short-circuited by the brushes. When the number of brushes is large, the parasitic brush currents affect the behaviour of the machine to a considerable extent. A detailed analysis of this phenomenon for a repulsion motor has been carried out by Morris.[32] In the stator-fed shunt motor, these currents increase the losses of the motor, produce cross-modulation of flux in the motor air-gap and develop additional torque. Kron[9] has suggested that the short-circuited coils may be replaced for analytical purposes by an additional layer of rotor winding below the main winding and having a different resistance and leakage reactance from the actual rotor layer. Hence, for accurate analysis, an extra mesh should be added in the actual and primitive systems of the machine while developing the tensor equations. Though in small machines the effect of cross-modulation of flux and of the additional torque may be neglected, any design-office calculation must take into consideration the extra no-load losses, particularly at high values of slip.

(2.7.3) The Slip-Independent Portion of Armature Leakage Reactance.

In the theory developed in earlier Sections, the motor-armature leakage reactance is taken as sx_{r1} (since the self-reactance of the motor secondary winding is taken as $sX_{r1} = sx_{r1} + k_m^2 sX_{m1}$). This assumes that the leakage reactance of the motor armature as measured at the brush terminals is a linear function of the slip. Arnold,[1] Richter[7] and Franklin[25] have all shown that the leakage reactance of an armature with a commutator winding, measured across the brush terminals, includes a part which is independent of the slip. Adkins and Gibbs[8] have suggested that this slip-independent portion is usually about one-fourth of the total standstill value. The presence of this portion can easily be taken into account by substituting $(sx'_{r1} + x''_{r1})$ in place of sx_{r1}, where x'_{r1} is the slip-dependent and x''_{r1} the slip-independent component of the armature leakage reactance.

(2.7.4) Variation of No-Load Losses with Speed.

In the general derivation of equations of performance, the losses in the magnetic circuit were neglected. It was stated later that the iron losses at the operating voltage could be separately added, or the mutual impedance Z_{om} could be considered to have a resistance component. Since the motor operates at practically constant voltage, the mutual-inductance component of Z_{om} can be considered as constant, but since the no-load electrical losses in the machine are a complicated function of speed, a fixed resistance in Z_{om} will not represent actual conditions. In order to take the effect of variation of no-load electrical losses with speed into consideration, the resistive and reactive components of Z_{11} can be considered to be in parallel, the reactive branch having a fixed value depending on the magnetizing current, and the resistive branch being expressed as an experimentally determined function of slip.

(2.7.5) Unbalance, Harmonics and Non-Sinusoidal Flux Density.

The theory has been developed for steady-state balanced operation, and the effect of unbalance, harmonics and non-sinusoidal flux-density distribution due to commutation is beyond the scope of the paper. In normal operation of the machine these factors always affect its behaviour to some extent, the effect being particularly noticeable near synchronous speed.

(3) CONCLUSIONS

The purpose of the paper was to evolve a comprehensive theory of the stator-fed polyphase shunt commutator motor. The ease with which tensor analysis can be used for developing equations of performance and establishing equivalent circuits has been demonstrated. Some tests have been carried out to verify the theory and have given fairly encouraging results.[33] It is intended to discuss the practical measurement of parameters and the comparison of test results with theoretically predicted behaviour in a later paper.

(4) ACKNOWLEDGMENTS

(5) BIBLIOGRAPHY

Books

(1) ARNOLD, E.: 'Die Wechselstromtechnik', Vol. V2 (Springer, 1912).

(2) OLLIVER, C. W.: 'The A.C. Commutator Motor' (Chapman and Hall, 1927).

(3) HUNTER-BROWN, P.: 'Carbon Brushes and Electrical Machines' (Morgan Crucible Co., 1939).

(4) KRON, G.: 'A Short Course on Tensor Analysis' (Wiley, 1942).

(5) TEAGO, F. J.: 'The Commutator Motor' (Methuen, 1946).

(6) HAYES, M. E.: 'Current-Collecting Brushes in Electrical Machines' (Pitman, 1947).

(7) RICHTER, R.: 'Elektrische Maschinen', Vol. V (Springer-Verlag, 1950).

(8) ADKINS, B., and GIBBS, W. J.: 'Polyphase Commutator Machines' (Cambridge University Press, 1951).

(9) KRON, G.: 'Equivalent Circuits of Electrical Machinery' (Wiley, and Chapman and Hall, 1951).

(10) GIBBS, W. J.: 'Tensors in Electrical Machine Theory' (Chapman and Hall, 1952).

(11) TAYLOR, E. O.: 'Performance and Design of A.C. Commutator Motors' (Pitman, to be published).

Papers

(12) RIDSDALE, J. L. D.: 'The Theory of the Three-Phase Variable Speed Shunt Commutator Motor', *Electrician*, 1921, **86**, pp. 448, 478, 515, 546.

(13) PARKER SMITH, S.: 'Single- and Three-Phase A.C. Commutator Motors with Series and Shunt Characteristics', *Journal I.E.E.*, 1922, **60**, p. 308.

(14) KOSTENKO, M., and SAWALISCHIN, P.: 'Experimentelle und Theoretische Kreisdiagramme der Mehrphasigen Kollektormaschinen nach der Methode des Allgemeinen Transformators', *Elektrotechnik und Maschinenbau*, 1931, **49**, p. 101.

(15) SCHWARZ, B.: 'Die Neure Entwicklung des Standergespeisten Drehstrom-Nebenschluss-Kollektormotors', *ibid.*, 1935, **53**, p. 85.

(16) COTTON, H.: 'The Action of the Commutator on Polyphase Commutator Motors', *Electrical Times*, 1939, **95**, pp. 857 and 911.

(17) 'Motor Developments', *Electrical Review*, 1939, **124**, p. 455.

(18) 'Three-Phase Variable Speed Commutator Motor', *Engineer*, 1939, **167**, p. 310.

(19) 'Improved A.C. Motors', *ibid.*, 1939, **167**, p. 412.

(20) COULTHARD, W. B.: 'A Generalized Equivalent Circuit in the Theory of Polyphase Commutator Motors', *Transactions of the American I.E.E.*, 1940, **60**, p. 423.

(21) NURNBERG, W.: 'Stander- und Laufergespeiste Drehstrom-Nebenschluss Motoren mit Stromwendern', *Elektrotechnische Zeitschrift*, 1941, **62**, p. 817.

(22) RICHTER, R.: 'Experimentelle Ermittlung des Zusatzlichen Drehmomentes der Mehrphasigen Standergespeisten Maschinen mit Stromwender', *Elektrotechnik und Maschinenbau*, 1943, **61**, p. 333.

(23) HOWINSON, K. C.: 'Three-Phase Commutator Motors', *Electrical Review*, 1945, **137**, p. 495.

(24) GIBBS, W. J.: 'The Theory of the Schrage Motor', *Journal I.E.E.*, 1946, **93**, Part II, p. 621.

(25) FRANKLIN, P. W.: 'A Study of the Three-Phase Armature with Six Adjustable Brushes', *Transactions of the American I.E.E.*, 1948, **67**, p. 197.

(26) ADKINS, B., and GIBBS, W. J.: 'Polyphase Commutator Machines', *Proceedings I.E.E.*, Paper No. 782 U, September, 1948 (**96**, Part II, p. 233).

(27) SCHWARZ, B.: 'The Stator-Fed A.C. Commutator Machine with Induction Regulator Control', *ibid.*, Paper No. 849 U, October, 1949 (**96**, Part II, p. 755).

(28) SCHUISKY, W.: 'The A.C. Shunt Motor and its Phase Compensation', *Elektrotechnische Zeitschrift*, 1949, **70**, p. 435.

(29) KLIMA, V.: 'Theory of Three-Phase Shunt-Wound Commutator Motors', *Electrotechnicky Obzor*, 1950, **39**, p. 260.

(30) MORRIS, D.: 'Some Tests of an Exact Practical Theory of the Transformer', *Proceedings I.E.E.*, Paper No. 902 S, February, 1950 (**97**, Part II, p. 17).

(31) PARTON, J. E.: 'Polyphase Commutator Motors', *Transactions of the Institution of Engineers and Shipbuilders in Scotland*, 1951–52, **95**, p. 492.

(32) MORRIS, D.: 'An Investigation of Parasitic Brush Currents in Single-Phase Commutator Machines, with Particular Reference to the Repulsion Motor', *Proceedings I.E.E.*, Monograph No. 67, June, 1953 (**100**, Part IV, p. 114).

(33) JHA, C. S.: 'Tensor Analysis of the Steady-State Behaviour of the Stator-Fed Polyphase Shunt Commutator Motor', Thesis for Fellowship, Heriot-Watt College, June, 1955.

(34) JHA, C. S.: 'Theory and Equivalent Circuits of the Double Induction Regulator', *Proceedings I.E.E.*, Monograph No. 197 U, September, 1956 (**104** C, p. 96).

THE CONDUCTIVITY OF OXIDE CATHODES

Part 1.—Potential Distribution

By G. H. METSON, M.C., Ph.D., M.Sc., B.Sc.(Eng.), Associate Member.

SUMMARY

The paper is the first of a series dealing experimentally with various aspects of the conductivity of oxide cathodes. The present Part starts with a brief survey of existing knowledge and then proceeds to examine the practical circumstances under which a non-uniformity of potential gradient can arise across the matrix under a modest continuous current loading. Using active-nickel cores to enclose a rectangular slab of barium-strontium oxide, it is found by probe experiments that non-uniformity of potential gradient occurs only in the presence of residual gas, and that the gradient tends to increase in those parts of the matrix likely to accumulate negative ions.

(1) INTRODUCTION

(1.1) Object of Work

The object of the present paper is to describe an experimental survey carried out on the general properties of oxide-cathode conductivity. The failure of receiving valves by deterioration of electron emission is accompanied by large changes in matrix conductivity, and it is thought that a greater knowledge of the nature of these changes may help towards a closer understanding of the life processes of such valves.

It is proposed in this Part of the paper to examine the distribution of potential across the oxide matrix under various conditions likely to be found in common receiving valves, the whole of the work being carried out at the conventional working temperature of 1 020° K. It is shown that in a well-activated cathode the potential gradient is constant across the matrix, but that the gradient may increase towards the outer surface as deactivation occurs.

In subsequent Parts certain other aspects of the electron-transfer mechanism will be examined, and the following schedule will give some idea of the general shape of the work:

(a) Influence of ionic movements consequent on electron flow.
(b) Distribution of power dissipation in matrix in the activated and deactivated states.
(c) Resistance to passage of free electrons through a cold matrix.
(d) Relative importance of the activation states of the two boundary surfaces of the matrix, etc.

As a consequence of the experiments a matrix model has been constructed and will be set out in the last Part.

(1.2) Survey of Published Information

Although literature on the resistivity of the oxide cathode is widespread, there is as yet no firm understanding of the mechanism of electron transfer through the oxide matrix. The bulk of effort seems to have expended itself on determining the temperature dependence of resistivity, ρ (or conductivity, σ), and in this direction there is general agreement that the relationship

$$\log \sigma \propto 1/T \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

holds over a limited range of absolute temperature. Perhaps the most significant advance in this direction is that of Loosjes and

Vink,[1] who have shown that a much wider range can be covered if two intersecting straight-line characteristics of the type given by eqn. (1) are used. The two different characteristics are explained as being due to two different types of electron transfer: in the low-temperature range by solid *n*-type semi-conduction, and in the higher-temperature range by the drift of an electron cloud through the interstices of the porous oxide matrix. The two mechanisms have become known as 'solid' and 'pore' conduction, and conventional valves working at 1 020° K are commonly supposed to rely mainly on pore conduction.

These views on the transfer mechanism have been opposed by Hannay, McNair and White,[2] who find that a cathode working in helium at a pressure of one atmosphere has the same conductivity as a similar cathode working in a vacuum. It is argued that such a pressure of helium filling the pores would inhibit the drift of the electron cloud and the conduction mechanism must therefore be a wholly semi-conducting one making exclusive use of the activated (BaSr)O lattice. Loosjes and Vink[1] rebut this argument on the grounds that the mean free path of an electron in helium at one atmosphere is about one micron, and since the probable pore dimension in the matrix is of this order, there should be no effect on pore conductivity. The general question of solid or pore conductivity at the normal working temperature of 1 020° K therefore seems to be still an open one.

A certain amount of attention has been directed towards the phenomenon of potential rise through the oxide matrix under current load. Loosjes and Vink[3] use a series of embedded probes at different distances from the metal core and find a progressive rise of potential through the matrix. With the aid of a movable anode and some inspired extrapolation they predict the surface potential and show that a steep increase is to be expected in the surface layers under high-current-density pulse conditions. This supposedly high gradient at the surface has also been used by such workers as Coombes[4] to explain the decay effects in high-current-density pulse work on oxide cathodes. Wright,[5] however, points out that such effects may be incidental to basic cathode action and may be due to extraneous causes, e.g. positive-ion bombardment and deactivation of the cathode surface consequent on residual gas ionization in the cathode-anode space. This view is backed up by some convincing evidence on diodes, which show no pulse decay effects when they have been well aged to clear up residual gas phenomena. It seems, then, that while a rise of potential across the matrix occurs under current load, there is no certainty that its gradient is constant.

The last point of interest concerns an observed proportionality between the conductivity of an oxide matrix and the temperature-limited electron emission that can be drawn from its exposed surface. This relationship has been widely observed and is not a subject of controversy.

(2) POTENTIAL DISTRIBUTION

(2.1) Experimental Valves

The basic experimental valve is designed with an eye to the provision of the necessary experimental facilities whilst main-
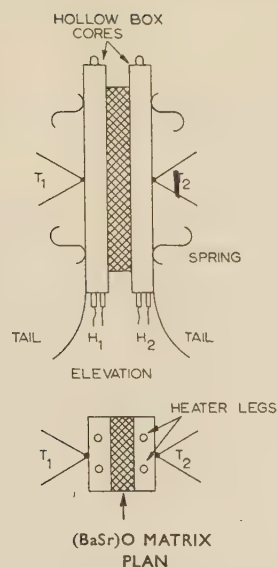
Fig. 1.—Arrangement of basic S-type assembly.

Fig. 2.—Relationship between core temperature and common heater voltage in S-type assembly.

...aining as far as possible a likeness to the cathode matrix of a common receiving valve. Fig. 1 shows the essential details of the arrangement, which will be described as an S-type assembly. Two rectangular box-like cores provided with insulated internal heaters are coated with oxide matrix and pressed together by light tungsten springs. The cores are prepared from a nickel alloy commonly employed in the manufacture of cathodes ('O' nickel) and contain the activating elements magnesium and silicon to the extent of about $0 \cdot 05\%$. The matrix is the conventional co-precipitated equimolar barium-strontium carbonate with a methacrylate binder, applied to the individual cores by spraying to an accurately known thickness and density. After breakdown of the carbonates to oxides, the density of the matrix is about $1 \cdot 0$ and the thickness between the core faces is about 150μ. The composition, form and density of the matrix are therefore similar to those used in common receiving valves, while the thickness is some 2–3 times greater. Power is led to and from the cores through the usual form of 'tail'. The whole arrangement is mounted on a standard valve pinch, sealed in a glass envelope and vacuum processed on a high-grade bench pump, the process schedule being shown in Table 1. A typical relationship between temperature of the cores and the common

Table 1

PROCESSING OF ACTIVE-NICKEL S-TYPE ASSEMBLY

(1) Seal to pump manifold and evacuate to manifold pressure of about 10^{-5} mm Hg.
(2) Bake at $400°$ C for 1 hour: pressure is then lower than 10^{-6} mm Hg.
(3) Degas getters.
(4) Decompose carbonates by passing current through both heaters in parallel.

Approximate temperature of cores °K	Time min
740	1
910	1
1 010	1
1 150	1
1 280–1 300	Until manifold pressure is below 2×10^{-6} mm Hg (approximately 3–5 min).

(5) Flash getters and seal from pump, with core temperature still 1 280–1 300° K.
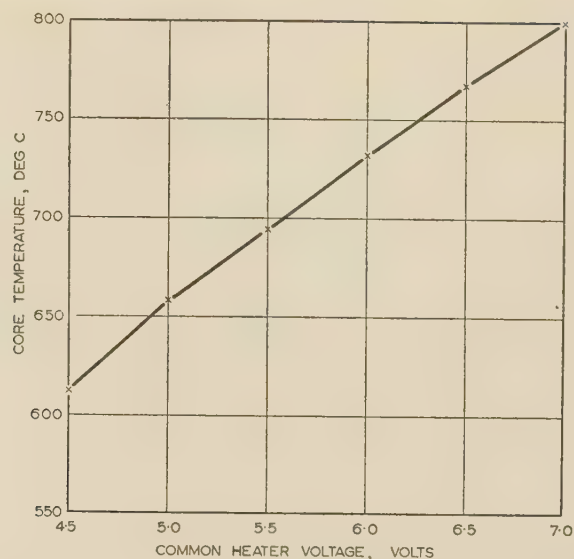
voltage applied to both heaters is given in Fig. 2 for future reference.

About a hundred of these basic valves were made during the course of the work, and they provided a vacuum 'diode' with metallic contact to both sides of a slab of (BaSr)O matrix approximately $9 \cdot 0$ mm long, $2 \cdot 5$ mm wide and 150μ thick. This basic S-type valve provides only limited experimental facilities, and the bulk of samples were, in fact, modified in detail to meet particular requirements. For example, in experiments involving the use of a temperature gradient across the matrix, the heaters in the two cores were brought out separately, so that the thermal power delivered to each core could be individually controlled. In such cases it was usually convenient to fit each core with its own thermocouple for purposes of temperature control. For measurements of potential within the matrix a fine (25μ) wire probe of pure platinum was included at the centre point of the matrix, and special arrangements were made to see that the lead-out wire of this probe was highly insulated from all other electrodes. Other special arrangements of the basic valve will be noted in the course of the work.

(2.2) Potential Measurement

The circuit arrangement for measurement of probe potential relative to either core is shown in Fig. 3. The heater circuits are supplied separately with d.c. power from large secondary cells to ensure an adequate short-term stability. The two thermocouples, T_1 and T_2, share a single switched instrument for measurement of core temperature. Power in the heater circuits is adjusted until $T_1 = T_2$ at the desired temperature with no current, I_A, in the main diode circuit. The driving of the main circuit is undertaken by a variable d.c. stabilized power-pack feeding through a 100-ohm ballast resistor. Power in the diode circuit and its resistance, R_d, are indicated by the milliammeter I_A and voltmeter V_A. The potential of the probe wire is measured by a high-quality potentiometer calibrated against a standard cell.

A potential measurement is carried out in the following manner: The current I_A is set accurately to some predetermined value (say 50 mA) and potentiometer measurements are taken between core 1 and the probe and between the probe and core 2. This operation results in two values, V_{P1} and V_{P2}. The current I_A is then reduced to zero and the measurements are repeated to
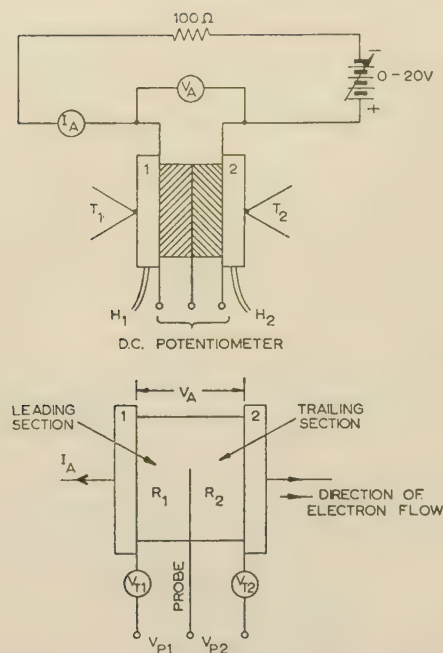
Fig. 3.—Circuit arrangements for probe potential measurement.

Fig. 4.—Invariance of resistance with current.

give the thermo-electric correction e.m.f.'s, V_{T1} and V_{T2}. Then, the potential across the leading section of the matrix is $V_{P1} - V_{T1}$, that across the trailing section of the matrix is $V_{P2} - V_{T2}$ and

$$V_A = (V_{P1} - V_{T1}) + V_{P2} - V_{T2} \quad . \quad . \quad . \quad (2)$$

should be closely realized in practical measurement.

A point of nomenclature must now be introduced. It will be observed in Fig. 3 that the electrons enter the system at core 1

Thus,

$$\left. \begin{aligned} R_1 &= \frac{V_{P1} - V_{T1}}{I_A} \\ R_2 &= \frac{V_{P2} - V_{T2}}{I_A} \\ R_d &= V_A/I_A \\ R_1 + R_2 &= R_d \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and

should be closely realized in practical measurement. Since the two sections of the matrix are in series to a common current, it follows that the potentials across the section are in similar ratio to the resistances of the sections. A typical test run, concerning the temperature characteristics of the two section resistances, is

Table 2

TYPICAL SERIES OF POTENTIOMETRIC MEASUREMENTS

V_H	V_A	I_A	V_{P1}	V_{P2}	V_{T1}	V_{T2}	R_1	R_2	$R_1 + R_2$	V_A/I_A
volts	volts	mA	volt	volts	volt	volt	ohms	ohms	ohms	ohms
4·0	2·130	50·0	0·640	1·521	−0·058	+0·067	13·95	29·10	43·05	42·6
4·5	1·590	50·0	0·437	1·172	−0·066	+0·076	9·92	21·90	31·82	31·8
5·0	1·162	50·0	0·337	0·834	−0·074	+0·086	8·22	15·00	23·22	23·2
5·5	0·870	50·0	0·291	0·579	−0·079	+0·094	7·40	9·70	17·11	17·4
6·0	0·680	50·0	0·217	0·471	−0·083	+0·102	6·00	7·38	13·38	13·6
6·5	0·600	50·0	0·179	0·426	−0·087	+0·110	5·32	6·33	11·65	12·0
7·0	0·530	50·0	0·143	0·397	−0·091	+0·119	4·68	5·56	10·24	10·6
7·5	0·481	50·0	0·115	0·377	−0·097	+0·129	4·24	4·97	9·21	9·64
8·0	0·454	50·0	0·095	0·366	−0·103	+0·139	3·96	4·54	8·50	9·08

and leave at core 2 after passing through the two halves of the matrix. That half of the matrix which is first traversed by an electron is termed the 'leading section', of resistance R_1, and the other half is termed the 'trailing section', of resistance R_2. If the direction of electron flow is reversed, the symbols in Fig. 3 change over on an image basis, i.e. an electron always enters the system at core 1, traverses the leading section before the trailing section and leaves at core 2.

At this stage a useful simplification can be introduced. It is shown in Fig. 4 that the matrix is almost ohmic in its voltage/current relationship over the current range 10–100 mA, and it is therefore possible to replace probe potentials depending on current I_A by the leading- and trailing-section resistances R_1 and R_2, which are independent of current.

set out in Table 2 to give some idea of the magnitude of experimental error to be expected in the measurement.

(2.3) Resistance/Time Characteristics

Perhaps the best point at which to open a discussion of the properties of the several resistances of the diode is by way of a life history or time characteristic. In new valves it is always found that the diode resistance, R_d, is high but decreases steadily with operating time to some stable value. The usual method of test is to set the two core temperatures to $1\,020^\circ\,$K and adjust I_A to 50 mA. Since R_d changes with time, I_A is maintained constant by variation of V_A. Regular potentiometric measurements of V_{P1}, V_{P2}, V_{T1} and V_{T2} are made to determine the time characteristics of R_1 and R_2.
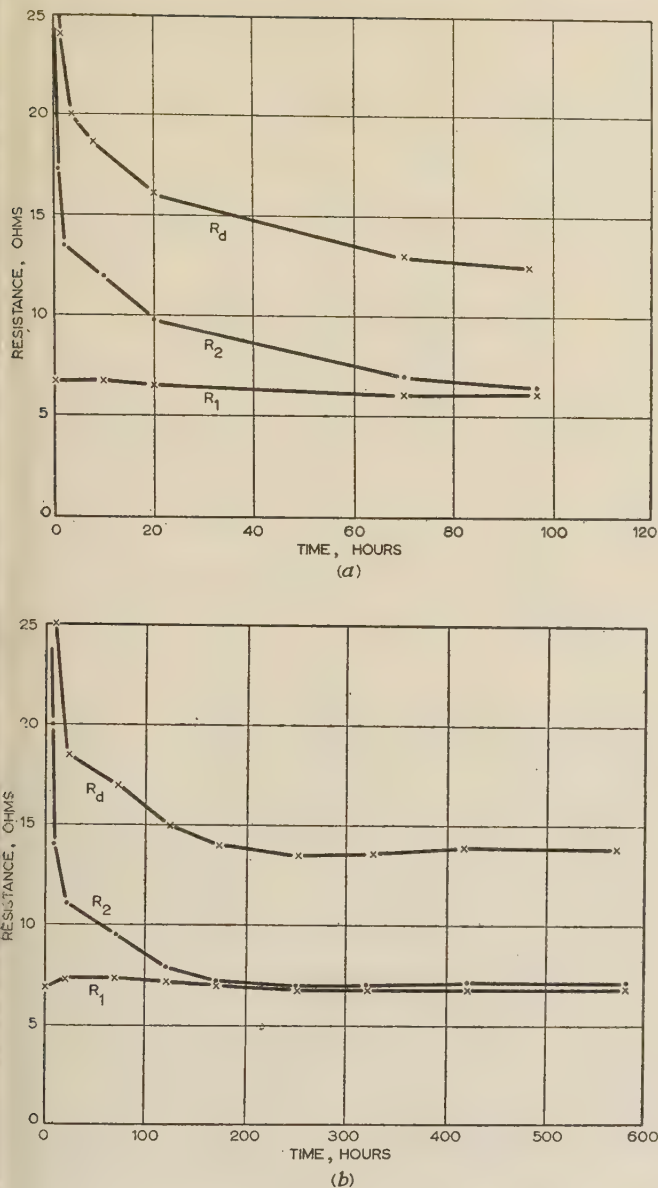
Fig. 5.—Time characteristics.
(a) 100 hours.
(b) 500 hours.

Time characteristics for R_1, R_2 and R_d for valve S-52 over a period of 100 hours are shown in Fig. 5(a), and those for S-51 over 500 hours in Fig. 5(b). These results can be regarded as being typical of all valves tested, and the following general conclusions are drawn.

(a) At $1\,020°\,\mathrm{K}$ R_d falls steadily from approximately 50 ohms to a constant value of about 14 ohms.

(b) Almost the whole of this fall in R_d is accounted for by decrement in R_2.

(c) Throughout the period of change of R_d and R_2 the leading resistance, R_1, remains substantially constant.

(d) When the stable low-resistance state of R_d has been reached the following condition holds:

$$R_1 = R_2 = \tfrac{1}{2}R_d \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The decrease in R_d with operating time will cause no surprise to valve engineers familiar with the properties of an oxide cathode with an active nickel core. The phenomenon, known as 'activa-

tion', is due either to a decrease in oxygen content or an increase of barium. What is perhaps surprising is that the whole of this betterment appears to occur in the trailing section of the matrix.

It has been shown that at zero time in the life of the S-type device the trailing resistance is always greater than the leading resistance, and this may seem peculiar in an arrangement which has physical symmetry. Two possibilities suggest themselves—either that the processing of the valve prior to test has given the device a directional property, or that polarization occurs on the first instant of electron flow. The first possibility can be eliminated, since the device is thermally activated on the pump and no current passes through the matrix until the time-characteristic measurement is actually started. The alternative can be readily tested by reversing the direction of electron flow during the period of decay of the diode resistance. Fig. 6 shows a time
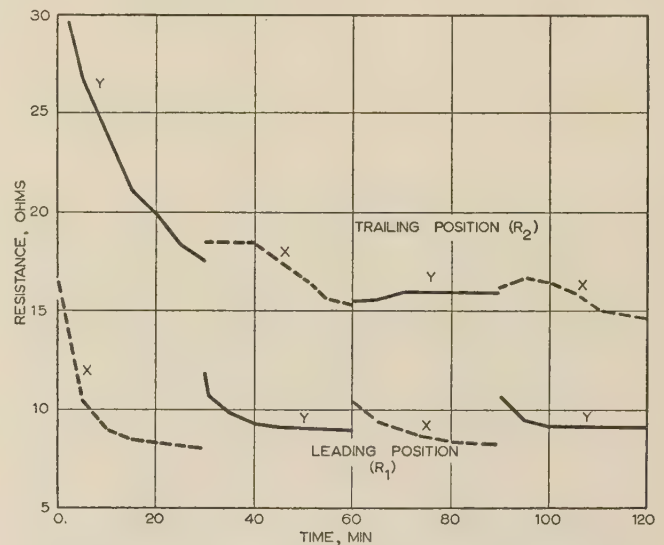
Fig. 6.—Effect of current reversal on resistance distribution.

characteristic of leading and trailing resistances carried out at a constant current of $50\,\mathrm{mA}$ and a cathode temperature of $1\,020°\,\mathrm{K}$. At zero time the leading and trailing sections of the device are designated X and Y respectively, and these letters are to be regarded as permanently fixed to the actual physical sections and independent of any subsequent change in direction of electron flow. For the first $30\,\mathrm{min}$ of life test the electron flow is in the X–Y direction with the X-section acting as leader and the Y-section as trailer; for the next $30\,\mathrm{min}$ the electron flow is reversed into the Y–X direction, with the Y-section moving into the leading position and the X-section moving into the trailing position; and so on in cyclic fashion every $30\,\mathrm{min}$ of the test.

The results in Fig. 6 show with some clarity that a physical section changes its activation state almost instantaneously with a change of direction of electron flow. The small difference in the leading-position resistances of the X- and Y-sections is possibly due to a minor deviation of the probe wire from the mid-point of the matrix.

(2.4) Influence of Oxygen Attack on Matrix

The influence of oxygen de-activation of the matrix on R_1, R_2 and R_d was examined in a standard S-type valve fitted with a controllable oxygen generator in the form of a tungsten-wire spiral coated with barium peroxide and heated by an externally regulated current. The valve was processed in the usual way, was sealed from the pump and its getter was flashed.

After 20 hours of operation with both cores at $1\,020°\,\mathrm{K}$ and a

current of 50 mA the valve had shown the normal decrement of R_d and was approaching the stable state of equality of R_1 and R_2, actual values being: $R_1 = 8 \cdot 1$ ohm, $R_2 = 10 \cdot 2$ ohm, and $R_2/R_1 = 1 \cdot 26$. The oxygen generator was now brought into action and adjusted for a period of 20 min to give a suitable rate of gas flow. During this period the current flow, I_A, was maintained constant at 50 mA by continuous adjustment of V_A, while regular potentiometric measurements were made on the probe wire. To avoid interference with I_A during the gas attack, the thermo-electric functions were not measured, but a comparison of the values at the beginning and end of the run showed that average values could be accepted without sensible error. Fig. 7
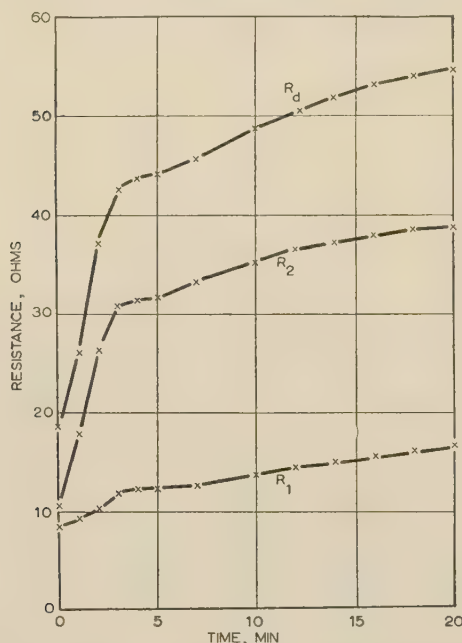
Fig. 7.—Influence of oxygen attack on resistance distribution.

shows the behaviour of R_1 and R_2 during the gas attack. At the end of the run the oxygen generator was cut off and the getter 'flash' was found to have shrunk in area to about one-half of its original size. It is of practical importance to leave some gettering power within the tube to work during the recovery phase.

Recovery of the system was carried through without change of the core temperatures ($1\,020°$ K) and with regular potentiometric measurement of the probe functions V_P and V_T. After 120 hours the whole system had regained the stable state with equality of the two section resistances and an overall resistance, R_d, of about 14 ohms, i.e.

$$R_1 = R_2 = \tfrac{1}{2}R_d = 7 \text{ ohms}$$

The recovery phase is shown in Fig. 8, and the ratio R_2/R_1 for both phases is plotted in Fig. 9. From these results the following conclusions can be drawn:

(a) During matrix de-activation the trailing resistance grows more rapidly than the leading resistance.

(b) During matrix activation the leading resistance falls more rapidly to the low stable level than the trailing resistance.

(c) There must be some significance in the fact that the overall resistance, for the constant matrix dimensions used, approaches, but never seems to go lower than, about 14 ohms at $1\,020°$ K.

(2.5) Correlation between Resistance and Total or Temperature-Limited Emission

That the fall of R_d coincides with a rise in total or temperature-limited emission of a well-processed (BaSr)O cathode on an
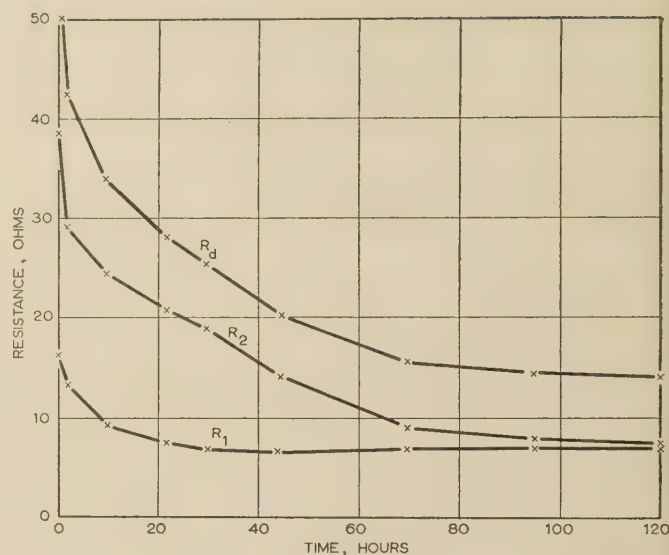
Fig. 8.—Reactivation after oxygen attack.
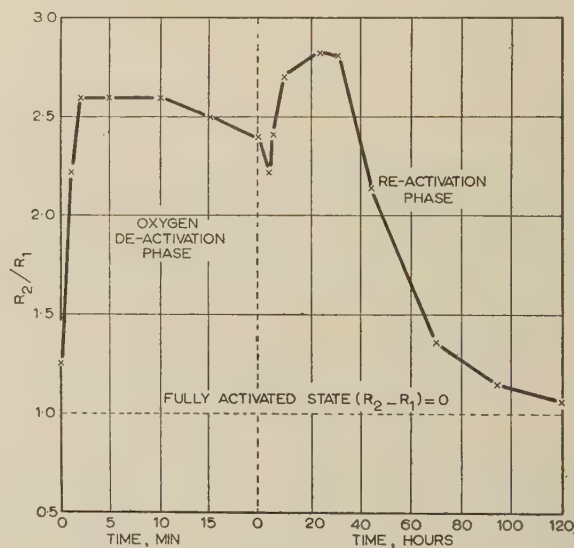
Fig. 9.—Time characteristic of R_2/R_1.

active-nickel core is shown in Fig. 10, the total-emission characteristic being taken as the mean of a group of conventional-type diodes with closely-spaced electrodes measured by the low-temperature ($700°$ K) technique.[6] The diodes have the same active-nickel core as the S-type valves and the oxide matrix is similar in constitution but only half the thickness.

The result shows a pronounced activation* effect in the conventional diodes, and it is suggested that this is due to the same cause as gives rise to the decay of R_d in the S-type valve.

(2.6) Elimination of Barium Orthosilicate Interface Layer as Cause of Resistance Unbalance

The S-type assembly uses the same type of core metal as common receiving valves which are known to develop a high-resistance

---

* The reason for the rise in total emission in active-core oxide-cathode diodes has previously[7] been considered as being due to an excess of metallic barium in the matrix over the optimum value for minimum work function. More recent experience has changed this view, and the emission rise is now believed to be due to a decrease in the gas content of the matrix.
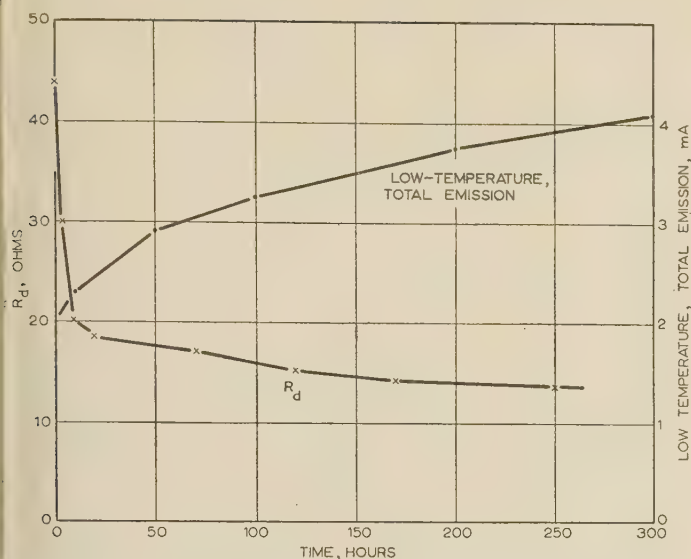
Fig. 10.—Correlation between changes of matrix resistance of S-type assembly and total emission (at 700° K) of conventional diode.

interface—supposedly due to a reaction between the silicon contaminant in the core metal and barium oxide:

$$Si + 4BaO \rightarrow Ba_2 SiO_4 + 2Ba \quad . \quad . \quad . \quad . \quad (5)$$

This barium orthosilicate by-product forms as a thin layer at the boundary between the core and the matrix. Owing to the presence of excess barium, the layer has a very low resistance (<1 ohm) at the beginning of life, but as time goes on and the barium becomes dissipated, the resistance rises to 100 ohms or more. Now a potential gradient across the matrix consequent on current flow results in an electrolytic action with migration in preferred directions of negative oxygen and positive barium ions. In the S-type assembly this action may conceivably result in the de-activation of the orthosilicate layer at the trailing core and further activation of the layer at the leading core. The whole directional phenomenon of resistance unbalance [$(R_2 - R_1) > 0$] could therefore be readily explained in terms of the interface layer. Fortunately, the question can be easily settled by removal of the silicon contaminant from the trailing core: if resistance unbalance disappears, the phenomenon is due to the orthosilicate action.

Two groups of experiments were undertaken to decide the question. In the first, assemblies were built with the trailing core prepared from pure electrolytic nickel containing less than one-tenth of the silicon possessed by the standard 'O' type nickel. This group behaved in the normal way, exhibiting an unbalance which approached zero after 100–200 hours' running time under the usual test condition. In the second group the trailing cores were prepared from very pure platinum; these, too, showed the normal unbalance reaction.

The conclusion reached from these tests is that de-activation of an orthosilicate interface layer makes no appreciable contribution to the resistance unbalance phenomena shown in Fig. 5.

(3) DISCUSSION OF RESULTS

The essential feature of the experimental valve used in the work is its dimensional symmetry about a probe wire buried in the centre of the (BaSr)O matrix. Other aspects of the valves are the use of active-nickel cores, which tend to increase the barium concentration in the oxide lattice, the avoidance of temperature gradients and the use of a matrix structure com-

parable in density and composition with that used in common valves. All work has been carried out at the conventional operating temperature of 1 020° K.

Under the conditions outlined above, the matrix has been found to have two distinct operating states, namely a transient state in which the diode resistance is high and a function of running time, and a state where it is low and independent of time. These states are identified with the degree of activation of the matrix, i.e. with its concentration of excess barium metal. If the overall resistance is analysed by potentiometric measurements across the two dimensionally similar sections of the matrix, two further resistances, R_1 and R_2, emerge, and the ratio R_2/R_1 is found to define the state of the matrix. Furthermore, the potentials across the sections are in similar ratio to the resistances and we can write,

Stable or activated state:

$$\frac{R_2}{R_1} = \frac{V_2}{V_1} = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Unstable or de-activated state:

$$\frac{R_2}{R_1} = \frac{V_2}{V_1} > 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

If the thickness of each section is d, the potential distribution across the matrix can be drawn for both states as a 3-point 'fix' as in Fig. 11(a). In the activated case the potential is a linear function of distance measured normally through the matrix with the gradient constant throughout; in the de-activated case the gradient is steeper in the trailing section than in the leading section. The distribution is settled in Fig. 11(a) from a 3-point 'fix', but a more likely form is visualized in Fig. 11(b).
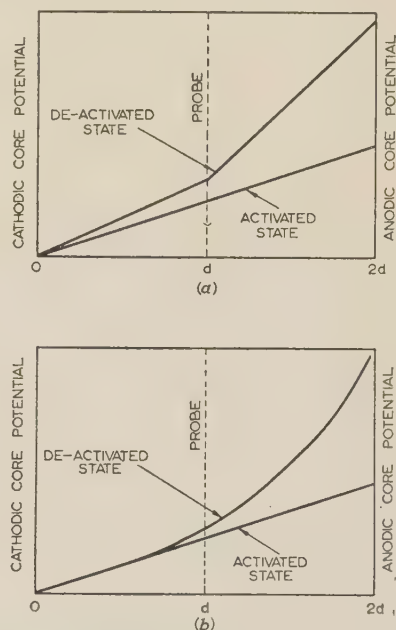
Fig. 11.—Potential distribution through matrix in activated and de-activated states.

(a) Linear distribution from 3-point measurement.
(b) More probable distribution.

An attempt must now be made to explain the discontinuity of resistance or potential gradient across the matrix in terms of matrix de-activation. This phenomenon of de-activation is taken as being synonymous with gas attack, and it involves the destruction by oxidation of a proportion of the excess barium metal on

which the matrix activity wholly depends. The supposed course of action is on the following lines: oxygen enters the vacuum pores of the leading and trailing section and becomes negatively ionized; under the influence of the potential gradient across the matrix the ions drift in the direction of the trailing core, about which they begin to congregate in atomic form after discharge; a concentration gradient of oxygen is therefore set up across the matrix, with the trailing section tending to accumulate a progressively higher proportion of the total quantity of gas within the matrix. The action of the gas within the pores is readily imagined. An atom or ion within a pore is continually rebounding from the (BaSr)O walls until it strikes by chance a free barium atom with which it combines. The number of journeys that it is likely to make before finding a free metal atom can be estimated from the known concentration of barium in the matrix and is about 10^4–10^5. The process of finding barium is therefore time-consuming and the rate of destruction within a pore will clearly be proportional to the number of gas atoms or ions present. The rate of de-activation of any particular layer within the matrix will therefore be directly proportional to the gas concentration within that layer, and that part of the matrix adjacent to the anodic or trailing core must be expected to suffer most.

Suppose the supply of gas is now shut off. As the last negative ion crosses the central plane of the matrix the leading section becomes gas-free and has only to make good its barium losses to recover its fully active or low-resistance state. These barium losses are relatively small, since by probability and symmetry, only half of the total gas entering the matrix goes into the leading section and then, owing to ionic migration, spends only a fraction of its life within the section. Barium losses in the trailing section are, of course, much heavier, since half the total gas entering the matrix spends its whole life, and the other half a fraction of its life, there. Moreover, the bulk of destruction will have occurred close to the anode, where matrix de-activation may well approach completion. The process of re-activation is dependent on the chemical nature of the active-nickel cores, which dispense magnesium vapour into the matrix.

Thus $$Mg + BaO \rightarrow MgO + Ba \quad . \quad . \quad . \quad (8)$$
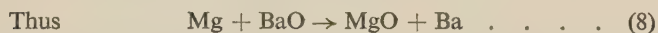
Since the cores have the same composition and temperature, the rate of barium generation is the same in both sections, and the rate of recovery of a particular section will therefore be a function of the degree of de-activation to which it has been previously subjected.

It will be appreciated that the essential step leading to the resistance-unbalance condition, $(R_2 - R_1) > 0$, is the conversion to the negatively-ionized state of gas atoms or molecules lying in the matrix pores. The assumption has been made in the paper that this conversion occurs as a form of contact phenomenon between a gas atom and the hot pore walls of the activated (BaSr)O matrix. Such a mechanism is not understood, but its existence is supported by the evidence of many writers[8,9] working with oxide cathodes in an imperfect vacuum.

(4) APPLICATION TO RECEIVING-VALVE CATHODES

It has been shown that, in the presence of oxygen or in an otherwise de-activated state, the S-type assembly develops a high resistance in the matrix adjacent to the electron exit boundary. The resistance of the layer is largely independent of current and is present at such moderate cathode loadings as $10 \, mA/cm^2$.

There is now a clear temptation to apply the result direct to the common receiving valve, and to assume that in the de-activated

state its cathode develops a high-resistance surface layer. The two cases, however, have one essential difference, namely that the electron exit boundary in the receiving valve is open to the escape of both electrons and negative ions of oxygen, whereas in the S-type valve only the electrons escape, leaving the discharged ions to congregate in, and de-activate, the trailing section of matrix. This free exit to negative ions therefore seems to preclude the possibility of surface-layer de-activation in receiving valves, but here again the obvious deduction proves wrong in practice. Consider, for example, the hypothetical case of an S-type assembly working in the unstable state, $(R_2 - R_1) > 0$, and with a trailing core of platinum, which can be regarded as being unreactive to oxygen. If the trailing core is now moved away from the matrix by a distance b which is small compared with the length and breadth of the matrix face, the assembly becomes a conventional vacuum diode with closely-spaced electrodes. Both electrons and negative ions now freely traverse the free surface of the matrix and fall on the anodic core, where the ions are discharged but not chemically absorbed. The oxygen atoms are thus free to move, and many of them—owing to the geometry of the system—will return to the matrix surface to de-activate the surface layer. The efficiency of the oxygen-return mechanism clearly increases as b decreases, until in the limit it becomes equal to unity. The S-type assembly is therefore the limiting case, with unity return factor, of the more general system typified by the common receiving valve, with a return factor always less than unity. This hypothesis concerning the development of high-resistance surface layers in receiving valves will be put to test in a subsequent Part, and a case will be made for its actual existence.

(6) REFERENCES

(1) LOOSJES, R., and VINK, H. J.: 'Conduction Mechanism of Oxide-Coated Cathodes', *Philips Research Reports*, 1949, **4**, p. 449.

(2) HANNAY, N. B., McNAIR, D., and WHITE, A. N.: 'Semi-Conducting Properties in Oxide Cathodes', *Journal of Applied Physics*, 1949, **20**, p. 669.

(3) LOOSJES, R., and VINK, H. J.: 'Distribution du potentiel dans la couche d'oxydes', *Le Vide*, 1950, **5**, p. 731.

(4) COOMBES, E. J.: 'The Pulsed Properties of Oxide Cathodes', *Journal of Applied Physics*, 1946, **17**, p. 647.

(5) WRIGHT, D. A.: 'D.C. and Pulsed Emission from Oxide Cathodes', *Proceedings of the Physical Society*, 1949, **62B**, p. 398.

(6) METSON, G. H., WAGENER, S., HOLMES, M. F., and CHILD, M. R.: 'Life of Oxide Cathodes in Modern Receiving Valves', *Proceedings I.E.E.*, Paper No. 1137 R, June, 1951 (**99**, Part III, p. 69).

(7) METSON, G. H.: 'A Study of the Long-Term Emission Behaviour of an Oxide-Cathode Valve', *ibid.*, Paper No. 1790 R, April, 1955 (**102 B**, p. 657).

(8) SLOANE, R. H., and WATT, C. S.: 'Emission of Negative Ions from Oxide-Coated Cathodes', *Proceedings of the Physical Society*, 1948, **61**, p. 217.

(9) VICK, F. A., and WALLEY, C. A.: 'Negative-Ion Emission from Oxide-Coated Cathodes', *ibid.*, 1954, **67**, p. 169.

SIMULATION OF THE TRANSIENT PERFORMANCE OF SYNCHRONOUS MACHINES ON AN A.C. NETWORK ANALYSER

By COLIN ADAMSON, M.Sc.(Eng.), Associate Member, and A. M. S. EL-SERAFI, B.Sc., Ph.D., Graduate.

SUMMARY

In order to make further progress with the solution of multi-machine transient problems, it is necessary to augment the facilities of network analysers, the most convenient of the analytical tools available to the power-system engineer. The paper describes a new simulator which enables the symmetrical a.c. transient component of current or voltage of a synchronous machine, subsequent to any 3-phase disturbance, to be obtained. The simulator is based on the duals of equivalent circuits for the direct and quadrature axes of synchronous machines, and is entirely electronic in its operation.

Experimental results are given and a comparison is drawn between these and the results obtained by calculation. It is concluded that such a simulator is an essential preliminary step towards a more accurate representation of machines and their associated apparatus, either individually or when interconnected in a power system.

LIST OF PRINCIPAL SYMBOLS

E = Induced e.m.f. in the synchronous machine.

E_q = Voltage behind synchronous quadrature-axis reactance.

V_t = Terminal voltage.

V_d, V_q = Voltages impressed on the direct- and quadrature-axis windings, respectively.

V_{d0}, V_{q0} = Initial values of the direct- and quadrature-axis voltages at short-circuit point.

V_f = Voltage impressed on the field winding.

E' = Voltage behind transient direct-axis reactance.

E_q' = Voltage behind transient impedance of the synchronous machine.

E'' = Voltage behind subtransient direct-axis reactance.

E_q'' = Voltage behind subtransient impedance of the synchronous machine.

V_R = Voltage behind the resistance of the synchronous machine.

V_{Rd}, V_{Rq} = Components of V_R in the direct and quadrature axes, respectively.

V_l = Voltage behind the leakage reactance of the machine.

V_{ld}, V_{lq} = Components of V_l in the direct and quadrature axes, respectively.

$G_{(p)}$ = Operator for voltage V_f

$$= \frac{(R_{kd} + X_{kd}P)X_{ad} - X_{ad}^2 p}{(R_{fd} + X_{fd}p)(R_{kd} + X_{kd}p) - X_{ad}^2 p^2}$$

I = Load current of the machine.

I_d, I_q = Components of I in the direct and quadrature axes, respectively.

R = Resistance of the synchronous machine.

R_e = External resistance connected to the machine.

R_{fd} = Field-winding resistance.

R_{kd} = Direct-axis damper-winding resistance.

R_{kq} = Quadrature-axis damper-winding resistance.

T_d' = Transient direct-axis time-constant =

$$= \frac{X_{fd} - \left(\frac{X_{ad}^2}{X_d + X_e}\right)}{R_{fd}}$$

T_d'' = Subtransient direct-axis time-constant

$$= \frac{(X_d + X_e)(X_{kd}X_{fd} - X_{ad}^2) - X_{kd}X_{ad}^2 + 2X_{ad}^3 - X_{fd}X_{ad}^2}{R_{kd}[(X_e + X_d)X_{fd} - X_{ad}^2]}$$

T_q'' = Subtransient quadrature-axis time-constant

$$= \frac{X_{kq} - \left(\frac{X_{aq}^2}{X_q + X_e}\right)}{R_{kq}}$$

T_a = Armature short-circuit time-constant

$$= \frac{2(X_d'' + X_e)(X_q'' + X_e)}{(R + R_e)(X_d'' + X_q'' + 2X_e)}$$

X_e = Reactance external to the machine.

X_d = Synchronous direct-axis reactance.

X_q = Synchronous quadrature-axis reactance.

X_l = Armature leakage reactance of the machine.

θ_0 = Angle of phase 'a' with respect to the direct-axis at the instant of short-circuit.

X_{ad} = Direct-axis magnetizing reactance.

X_{aq} = Quadrature-axis magnetizing reactance.

X_{kd} = Direct-axis damper-winding reactance.

X_{kq} = Quadrature-axis damper-winding reactance.

X_{fd} = Field-winding reactance.

$X_{d(p)}$ = Direct-axis operational reactance

$$= X_d - \frac{(R_{kd} + X_{kd}p)X_{ad}^2 p + (R_{fd} + X_{fd}p)X_{ad}^2 p - 2X_{ad}^3 p^2}{(R_{fd} + X_{fd}p)(R_{kd} + X_{kd}p) - X_{ad}^2 p^2}$$

$X_{q(p)}$ = Quadrature-axis operational reactance

$$= X_q - \frac{X_{aq}^2 p}{R_{kq} + X_{kq}p}$$

X_d'' = Subtransient direct-axis reactance

$$= X_d - \frac{X_{kd}X_{ad}^2 + X_{fd}X_{ad}^2 - 2X_{ad}^3}{X_{fd}X_{kd} - X_{ad}^2}$$

X_d = Transient direct-axis reactance = $X_d - \frac{X_{ad}^2}{X_{fd}}$.

X_q'' = Subtransient quadrature-axis reactance

$$= X_q - \frac{X_{aq}^2}{X_{kq}}$$

$p\theta$ = Angular velocity of rotation.

X_m = Reactance determining the asymmetrical component

$$= \frac{2(X_d'' + X_e)(X_q'' + X_e)}{X_d' + X_q'' + 2X_e}$$

Correspondence on Monographs is invited for consideration with a view to publication.

Mr. Adamson and Dr. El-Serafi are in the Power Systems Laboratory, Faculty of Technology, University of Manchester, Dr. El-Serafi being on study leave from Cairo University.

X_n = Reactance determining the second-harmonic component

$$= \frac{2(X_d'' + X_e)(X_q'' + X_e)}{X_q'' - X_d''}$$

(1) INTRODUCTION

Synchronous generators, as sources of electrical energy in power systems, are of prime importance, and there has been interest recently in improved ways of predetermining their performance. The fundamental theory of this class of apparatus is substantially complete. Convenient analytical methods for the solution of transient problems of machines connected to power systems do not, however, exist; attention has consequently been paid to improved methods of simulation of synchronous machines. Although much work has been published which deals with the simulation of inertia and damping, little progress has been made towards simulation of the purely electrical characteristics of the machines, automatic voltage-regulator action and saturation.

The theoretical basis and construction of a simulator for the fundamental-frequency transient performance of the synchronous machine is described in the paper, together with the results of the simulation of a generator under short-circuit conditions.

(2) PROBLEMS OF SYNCHRONOUS MACHINES

(2.1) Problems of Synchronous Stability

Stability is concerned with the ability of the machine to remain in synchronous operation under abnormal system conditions; it is usual[1] to neglect the effect of regulating systems, and the machine in the transient state is represented as a constant voltage behind a constant impedance. This gives an acceptable solution only for the first swing of the machine subsequent to an external disturbance; however, consideration of this problem should not be restricted to the first swing alone if high-speed regulator effects are to be taken into account.

(2.2) Problems of Electrical Characteristics

The problems of electrical characteristics are concerned with the predetermination of electrical quantities, i.e. current, voltage and electrical torque, from the known data and parameters of the machine and, where necessary, the system to which it is connected. Information relating to these quantities is needed for fault studies and for protective-gear selection. Unlike the case of stability, with problems of electrical characteristics it is customary to assume that the machine speed is constant; this has a sound foundation only for small rotor angular changes. With voltage-regulator action, the terminal voltage is appreciably affected by rotor angle.

The paper is primarily concerned with the problems of electrical characteristics rather than synchronous stability.

(3) THEORETICAL APPROACH TO THE ELECTRICAL TRANSIENT CHARACTERISTICS

(3.1) The Two-Reaction Theory[2-6]

In synchronous machines, most self and mutual inductances between stator and rotor circuits are periodic functions of rotor angular position. Thus the characteristics of a synchronous machine may be expressed by simultaneous differential equations, most of whose coefficients are periodic functions of rotor angle. The 2-reaction theory was introduced to overcome this difficulty by replacing the three phases of the armature winding by two equivalent, but fictitious, stationary windings, one in the pole axis (direct axis) and the other in the interpole axis (quadrature axis). In consequence, the windings on each axis are mutually coupled as on a transformer but differ owing to the addition of voltage due to rotation. The schematic arrangement of windings of the 2-reaction theory is illustrated in Fig. 1.
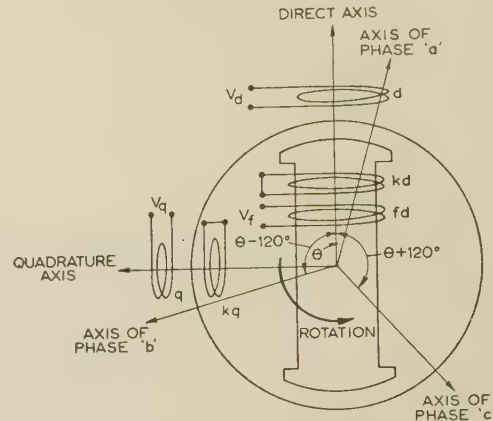
Fig. 1.—Ideal synchronous machine.

The general equation of the synchronous machine is then

$$\begin{bmatrix} V_d - G_{(p)}pV_f \\ V_q - G_{(p)}p\theta V_f \end{bmatrix} = \\ \begin{bmatrix} -(R+R_e)-p[X_e+X_{d(p)}] & [X_e+X_{q(p)}]p\theta \\ -[X_e+X_{d(p)}]p\theta & -(R+R_e)-p[X_e+X_{q(p)}] \end{bmatrix} \begin{bmatrix} I_d \\ I_q \end{bmatrix} \quad \quad (1)$$

If the machine speed is considered to be constant and there is no voltage-regulator action, the solution of eqn. (1) is straightforward. The solution is sinusoidal in form but with the amplitude varying with time; the wave may be displaced in phase from the reference axis. In the case of multi-machine problems, analytical solution of the equations corresponding to eqn. (1) is formidable, since, in addition to the expressions relating to the direct and quadrature axes of each machine, there are the transformation equations between the axes. Solution of such problems on a digital computer is, however, conceivable.

Considerable simplification in the solution of eqn. (1) can be obtained if only the fundamental-frequency component of the transient is required. This can be done by neglecting the terms $-(R_e + R) - p[X_e + X_{d(p)}]$ and $-(R_e + R) - p[X_e + X_{q(p)}]$, which is equivalent to ignoring the interaction between the direct and quadrature axes. (This is discussed in Section 11.1.)

Such problems may also be solved on a steady-state basis, and instantaneous values can be obtained by using, as appropriate, the subtransient, transient or synchronous reactances for the commencement of the subtransient, transient and steady-state periods, respectively. The voltage behind these reactances is dictated by the theorem of constant flux linkages;[7] thereafter, by the use of appropriate time-constants, it is possible to determine the conditions at any time. The vector diagrams of the synchronous machine in the steady state, and in the initial subtransient and transient states but on a steady-state basis, are shown in Fig. 2. The maximum possible d.c. component can be found by subtracting vectorially the initial load conditions from the initial instantaneous current in the subtransient period, and multiplying the resultant by $\sqrt{2}$. With the appropriate value of the time-constant and instant of short-circuit, the d.c. value at any other time can also be found. It is very difficult

QUADRATURE AXIS

DIRECT AXIS

Fig. 2.—Vector diagram of the synchronous machine.

linear phenomena are to be included; for the solution of such problems non-manual methods of computation are necessary. The methods available are simulation by small machines in miniature system, direct analogy (network analysers), mathematical analogue (differential analyser), or the use of a digital computer.

(4.1) Scale Models

Robert[14] has described the use of rotating machines of low power (1–10 kVA), which produce some of the same characteristics as large ones. Although this method of simulation is the most exact to date, it is complicated. It is not easy for some of the parameters, e.g. saliency, saturation and damping, to be adjusted over a wide range, and adjustment may take a long time. The number of components required to represent a given system is large, and the small rotating machines, under unbalanced conditions, differ from their prototypes mainly because the zero-sequence impedance is dependent on the design of the stator itself.

(4.2) Analogue Computers

For the solution of the equations of a synchronous machine, differential analysers have been found useful,[15, 16] but with complicated systems with many sets of differential equations, the amount of equipment required is formidable.

(4.3) Network Analysers

The network analyser is the computing device most readily available to electrical-power engineers. Also, it is needed for representing the extensive static network of the power system even when the devices mentioned in Sections 4.1 and 4.2 are used to represent the generator. There are several different ways of using a network analyser.

(4.3.1) Initial Instantaneous Quantities.

One method has been considered by the authors in an earlier paper.[17] Initial values, appropriate to the reactance and voltage-behind-the-reactance selected, are obtained; thereafter time-constants must be used for further elaboration of the solution.

(4.3.2) Coupled-Network Method.

A 2-reaction representation of a synchronous machine is possible through the reciprocal and non-reciprocal couplings of inductors and resistors. Such an arrangement may be set up on a network analyser for solution, the non-reciprocal coupling being represented by electronic amplifiers.

The coupled-network method is useful only for solving for single machines, since with multi-machine problems it is difficult to represent the transformation from one machine to another on the network analyser. The method may be useful if, for a given machine, there is interest in the electrical quantities in the different windings on the rotor.

(4.3.3) Adkins's Equivalent Circuits[6]

Adkins has developed equivalent circuits, shown in Fig. 3A, to determine the alternating components of the short-circuit current in the two axes separately. Unfortunately, these equivalent circuits are ideal since the inductances are loss-less and it is thus impossible to set them up directly on network analysers. They are, however, the basis of the simulator described in Section 5.

(4.3.4) Steady-State and Hunting Equivalent Circuits.

Kron[18] and Ku[12, 13] have established equivalent circuits resulting from the rotating-field theory to represent synchronous machines and their conditions of hunting (small oscillations superimposed upon the steady-state rotation). These can be

to find the values of time-constants for complicated cases, and the method appears to be suitable only when initial instantaneous values alone are required. This method can be extended for an approximate solution of the fundamental-frequency component under unbalanced conditions.

(3.2) α, β, 0 Components[8-10]

α, β, 0 components are useful in dealing with the unbalanced conditions of synchronous machines. These components are equivalent to the replacement of the three phases on the machine by two phases and a fictitious zero-sequence winding fixed with reference to the stator; this results in linear differential equations, the coefficients of which are variable. The solution for unbalanced conditions is not sinusoidal and may contain a large percentage of harmonics.

(3.3) The Rotating-Field Theory[11-13]

The rotating-field theory is an important alternative to the 2-reaction theory, and has recently been reintroduced by Ku. In it, the direct and quadrature axes are replaced by two axes, one rotating forward in the same direction as the rotor and the other rotating backward. Physical representation of this theory is difficult and computation is laborious.

4) USE OF COMPUTING MACHINES AND ANALYSERS

The theories quoted are extensible rarely to multi-machine problems, and never if the action of regulating systems and non-

Fig. 3A.—Equivalent circuits for alternating component of current after a sudden change.

set up on the network analyser, and interconnection between the equivalent circuits of any number and type of rotating machines and transmission lines may be made. For transient problems, the constant-flux-linkage theorem can be used to solve for the initial instantaneous values and, with appropriate time-constants, values at any other time can again be calculated. Once again, the time-constants are difficult to estimate in complicated cases.

(5) DIRECT SIMULATION OF TRANSIENT CONDITIONS ON A NETWORK ANALYSER

In the view of the authors, the most fruitful approach to transient analysis is through the network analyser and the development of economical apparatus for the simulation of generators in the transient state, thereby avoiding the formulation of differential equations and permitting considerable use of existing apparatus.

In its present form, the analogue simulates the fundamental-frequency component of the electrical transient under balanced

conditions, and may be used as an ordinary generator unit o[...] the network analyser. There is no theoretical limitation to t[...] number of machines being studied in the system, and it is thoug[...] that its principle is well suited to an easy solution of the voltag[...] regulator problem and to more unusual problems such [...] transient stability after the first swing; in this latter case, [...] would be necessary to associate the apparatus with automat[...] phase-modulating equipment such as that described in a[...] earlier paper.[19]

(5.1) Principle of the Simulator

Equivalent circuits containing lumped impedances are used t[...] represent the 2-axis equations of the synchronous machine. [...] representation such as this is not possible with the alternativ[...] rotating-field theory. The equivalent circuits are shown i[...] Fig. 3B. If the components of the voltage behind the resistanc[...] of the machine in the two axes have been applied as shown, th[...] solution gives the axis components of the fundamental frequenc[...] as proved in Section 11.2.

It is interesting to note that these equivalent circuits are th[...] dual of those of Adkins shown in Fig. 3A; they can be derive[...] therefrom by dividing all the applied voltages and impedances i[...] Adkins's equivalent circuits by the operator p.

(5.2) Essential Features of the Simulator

The simulator depends on the representation of a synchronou[...] machine by a varying voltage behind a constant reactance. Th[...] equivalent circuits in Fig. 3B are used by applying to thei[...] terminals the components of voltage behind the resistance [...] the machine, V_R. A phase-conscious rectifier followed by [...] low-pass filter and a d.c. amplifier is used to generate the com[...] ponents of V_R in each of the two axes, i.e. V_{Rd} and V_{Rq} ar[...] both d.c. signals; a suitable d.c. signal supplied through a d.[...]
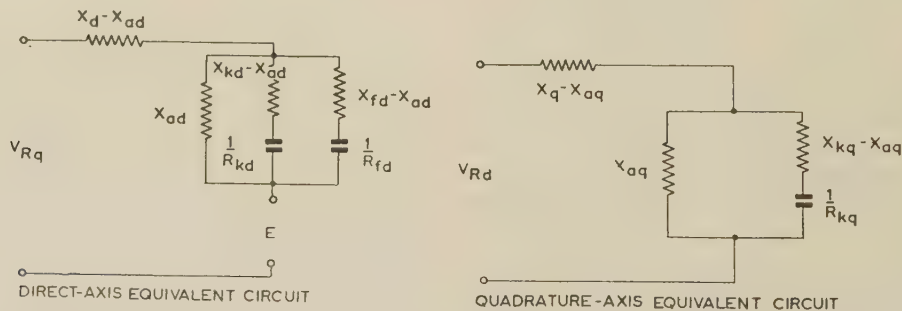
Fig. 3B.—Equivalent circuits for direct and quadrature axis of synchronous machine.
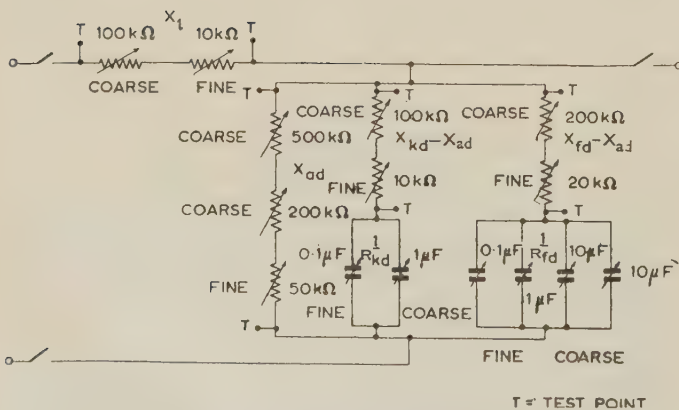
Fig. 3C.—Design details of direct-axis equivalent circuit.
Variable capacitors can each be switched in 10 steps.
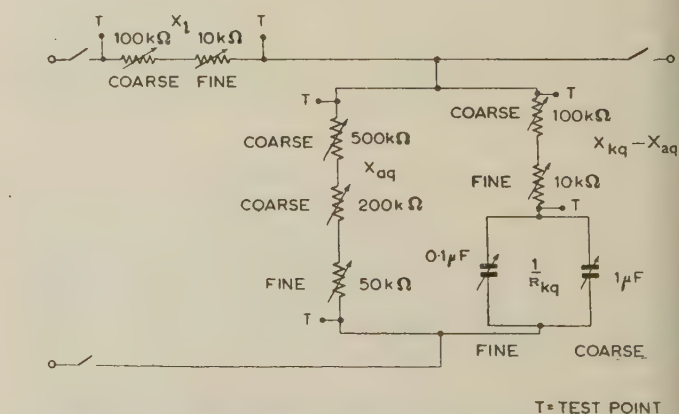
Fig. 3D.—Design details of quadrature-axis equivalent circuit.
Variable capacitors can each be switched in 10 steps.
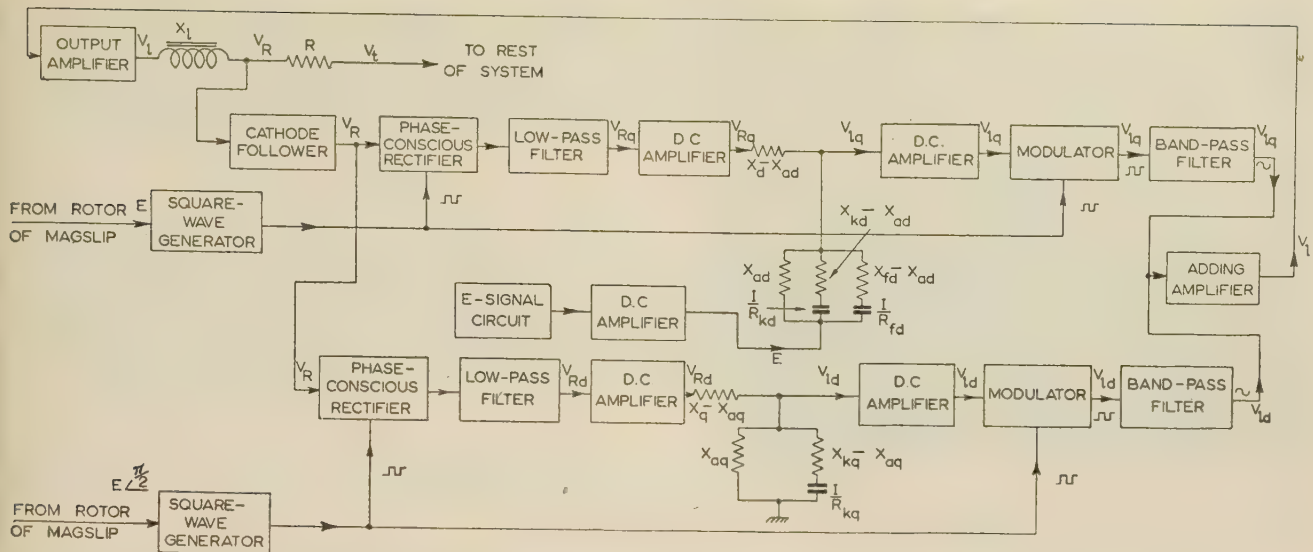
Fig. 4.—Simulation of transient characteristic of synchronous machine.

amplifier is used to represent the e.m.f., E. These voltages are applied to the equivalent circuits in the simulator, which consists of two feedback loops, one for the direct axis and the other for the quadrature axis, as shown in Fig. 4; this is equivalent to the transformation of the phase quantities on the machine to the axis quantities. The synchronous machine is represented on the network analyser by the leakage reactance and the voltage behind it, the latter obtained by modulating the two voltages V_{ld} and V_{lq} taken from the equivalent circuits. The modulated signals are filtered through band-pass filters, added, and fed back to the network analyser: this is equivalent to the transformation of the axis quantities back again into phase quantities. For any balanced disturbance in the power system, V_{ld} and V_{lq} will vary according to the transient equation of the synchronous machine, i.e. eqn. (11). It should be noted that the simulator also represents the synchronous machine under steady-state conditions.

(5.3) Time Scaling

By analysing the equivalent circuits of the two axes, the fundamental-frequency components of the transients in both the axes will result. These components can be transformed to phase quantities by the process of modulation in the simulator, the resultant being an a.c. wave which changes in magnitude with the axis quantities. It is possible for the time scale of the a.c. frequency to differ from that of the transient, as there is no relation between them. The frequency of the a.c. signal is a matter of the design of the network analyser, and the time scale of the transient is dependent mainly on the transient response of the various units of the simulator. The components with the longest rise-time are the filters, mentioned in Section 6.3; the reduction of this rise-time has been one of the most difficult design problems of the simulator. A scale factor of unity was found suitable, i.e. the per-unit time corresponds to 50 c/s.

Thus, $1 \text{ p.u. time} = \dfrac{1}{100\pi} \text{ sec.}$

The per-unit values of the components of the equivalent circuits were chosen as follows:

1 p.u. capacitance, i.e. 1 p.u. $\dfrac{1}{R} = 0.01\,\mu\text{F}$.

1 p.u. resistance, i.e. 1 p.u. $X = 318.5$ kilohms.

(6) DESIGN DETAILS OF THE SIMULATOR

(6.1) Phase-Conscious Rectifier and Modulator Circuits

The purpose of the phase-conscious rectifiers is to give d.c. outputs proportional to V_{Rd} and V_{Rq}, respectively; the operation depends on the design of a balanced rectifier modulator. The switching signals of the modulator are square waves, one in phase with the signal E and the other 90° out of phase with it, and supplied from two square-wave generators. Should the square waves be applied directly to the modulator, a large value of capacitance, of the order of $1\,\mu\text{F}$, would be required, with detriment to the transient response subsequent to any change in the a.c. signal. This difficulty may be overcome, and small values of capacitance can be used, if the square wave is applied by means of an r.f. carrier; this device was adopted, the grid of an r.f. oscillator being modulated by the square wave and the resulting output used for switching purposes, as shown in Fig. 5.
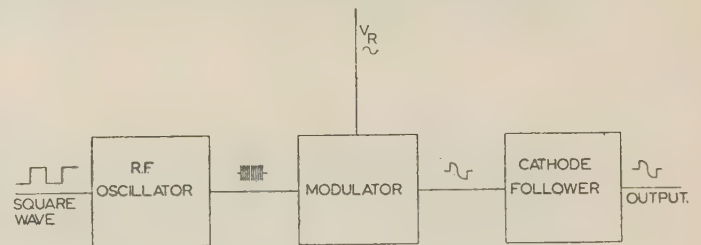
Fig. 5.—Block diagram of modulator arrangements.

Thus the input signal, V_R, which is applied through a cathode follower, is switched, and the output contains a d.c. signal proportional to V_{Rd} and V_{Rq}. The relevant waveshapes are shown in the first two sets of oscillograms of Fig. 6.

The same circuit design is used for modulation of the d.c. signals V_{ld} and V_{lq}, and the output waveforms are shown as the third set of oscillograms in Fig. 6.

(6.2) The Square-Wave Generators

The inputs to the square-wave generators are taken from two windings at right angles on the rotor of a magslip. The two required square waves are produced, one corresponding to the e.m.f. E, the other being 90° out of phase. The design is similar to that described in a previous paper.[17]

$\theta = 0°$ $\theta = 30°$ $\theta = 60°$ $\theta = 90°$
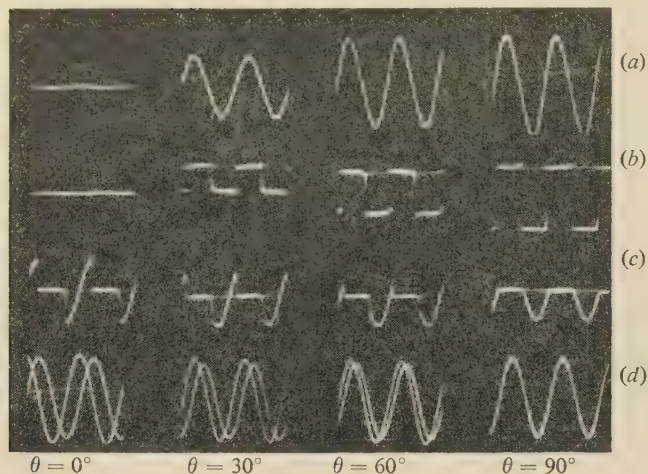
Fig. 6.—Oscillograms at different points on the feedback loop.

(a) Both the input and the signal, E.
(b) Output from the phase-conscious rectifier.
(c) Output from the modulator.
(d) Output from the feedback loop.

(6.3) Filters

Two types of filter are used. The low-pass filter is used to filter the modulated wave from the phase-conscious rectifier so as to leave only the d.c. components, which are proportional to V_{Rd} and V_{Rq}. The band-pass filter is used to filter the modulated wave from the modulator in order to separate the fundamental-frequency component. These filters require careful design since they have the longest time-response and greatly affect the accuracy. They are of standard form; it was found that the low-pass filter had a build-up time of about 1 millisec (0·314 p.u.), whilst the band-pass filter had 0·7 millisec (0·2 p.u.) build-up time.

(6.5) Adding Stage

In this stage the two signals V_{ld} and V_{lq} are added to give th[e] voltage behind the leakage reactance, V_l. This output is appli[ed] to the network analyser by a network-analyser output amplifi[er]

(6.6) The Design of the Equivalent Circuits

The equivalent circuits are built from commercial variab[le] resistances which can be adjusted by a standard bridge. Th[e] capacitance elements are decade units connected in parall[el] The magnitudes of the equivalent-circuit elements are shown [in] Figs. 3c and 3d.

(7) PERFORMANCE AND RESULTS

(7.1) Steady-State Performance

The overall steady-state performance of one of the feedba[ck] loops, with the other disconnected, shows that the loops a[re] linear within $\pm 6\%$ over the working range.

(7.2) Transient Performance

With reference to Fig. 3b, the equivalent circuits were set for [a] typical synchronous machine, the constants being as follows:

$X_d = 1·15$ p.u.	$X_q = 0·75$ p.u.
$X_{ad} = 1·0$ p.u.	$X_{aq} = 0·6$ p.u.
$X_{kd} = 1·1$ p.u.	$X_{kq} = 0·7$ p.u.
$X_{fd} = 1·25$ p.u.	$R_{kq} = 0·04$ p.u.
$R_{kd} = 0·03$ p.u.	
$R_{fd} = 0·001$ p.u.	

An input to the feedback loop, V_R (= 1 p.u.), was switched o[n] and off. The phase angle, with respect to E, was maintaine[d] zero in one case so that only the direct-axis equivalent circu[it] was energized; in the other case, the phase angle was maintaine[d]

(a)

(b)

Fig. 7.—Oscillograms of output voltage.

(a) From the quadrature-axis equivalent circuit.
(b) From the direct-axis equivalent circuit.

(6.4) D.C. Amplifiers

As shown in Fig. 4, the main purpose of these is to apply the signals to the equivalent circuits with low output impedance, and to isolate the equivalent circuits from the input of the modulators. Each consists of a Miller-compensator amplifying stage followed by a cathode-follower stage. Feedback is applied overall and the drift is better than 10 millivolt. The output impedance is about 60 ohms, i.e. about 0·02% of the per-unit resistance of the equivalent circuits.

at 90° out of phase so that the quadrature axis alone wa[s] energized.

The resulting input voltage to the equivalent circuits wa[s] found to have a reasonably vertical side and no ripples. Th[e] output from equivalent circuits [Figs. 7(a) and 7(b)] together wit[h] the output from the feedback loops [Figs. 8(a) and 8(b)] wer[e] analysed; the deviation from the theoretical calculation wa[s] $\pm 7\%$ in general, and at some points $\pm 10\%$. This includes th[e] errors involved in the analysis of the oscillograms.
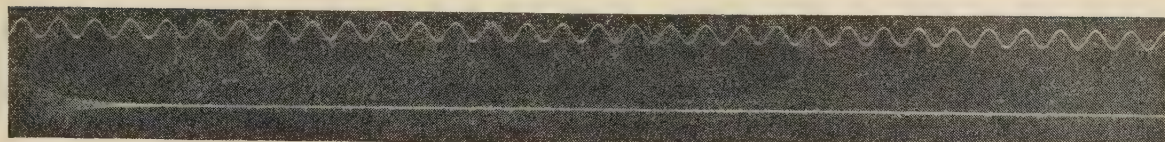
(a)
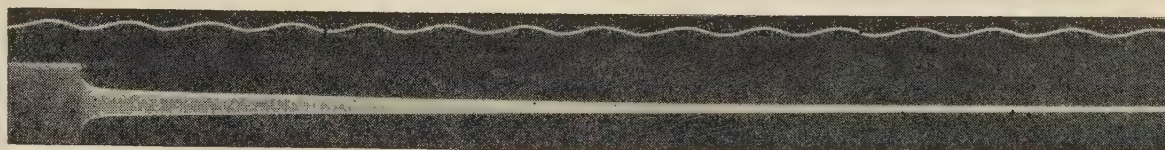
(b)

Fig. 8.—Oscillogram of the output for a step-function input.
(a) From the quadrature-axis feedback loop.
(b) From the direct-axis feedback loop.
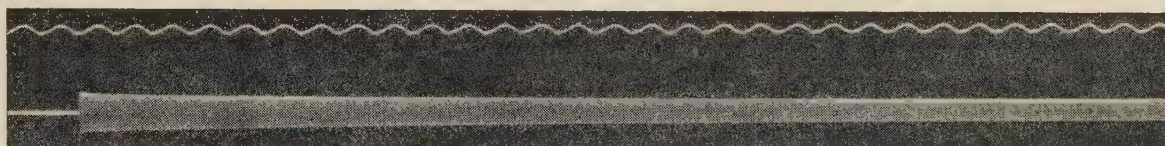
Fig. 9.—Oscillogram of the alternating component of the short-circuit current from the no-load condition.

(7.3) Study of a Synchronous-Machine Problem

As a further check on the performance of the simulator and to demonstrate its principle, a simple problem, which could also be calculated, was considered. The problem is that of a synchronous machine connected to a busbar through a 1 p.u. external reactance. The constants of the synchronous machine are the same as those mentioned in Section 7.2. The e.m.f., E, of the synchronous machine and voltage of the busbar were kept constant at 1 p.u. A short-circuit was applied at the mid-point of the external reactance, the initial condition being no-load. Fig. 9 shows the oscillogram of the alternating component of the short-circuit current; Fig. 10 shows in detail the shape of the envelope of this short-circuit current, theoretical and simulator results being compared. The calculated equation for the envelope of the short-circuit current is

$$i_{envelope} = 0 \cdot 309\varepsilon^{-t/8 \cdot 42} + 0 \cdot 807\varepsilon^{-t/644} + 0 \cdot 857 \text{ p.u.}$$

The equation of the same envelope, but derived from the results of the simulator, is

$$i_{envelope} = 0 \cdot 34\varepsilon^{-t/8 \cdot 5} + 0 \cdot 73\varepsilon^{-t/665} + 0 \cdot 87 \text{ p.u.}$$

It may be seen, from a comparison of these two expressions, that the error in the time-constant is within $\pm 3 \cdot 5\%$, for the initial values within $\pm 10\%$, and for other points on the curve within $\pm 5\%$. Again, this includes errors in the analysis of the oscillograms.

(7.4) Discussion of the Results and Accuracy

The main sources of error are:

(a) Inherent errors in the phase-sensitive rectifiers and modulators, estimated to be $\pm 3\%$ in each case.
(b) Drift in the zero settings of the individual units of the apparatus, especially in the d.c. amplifiers.
(c) The accuracy in setting up the problem; this depends on the accuracy of the network analyser, which was estimated to be $\pm 5\%$.
(d) The error in the analysis of the oscillograms.
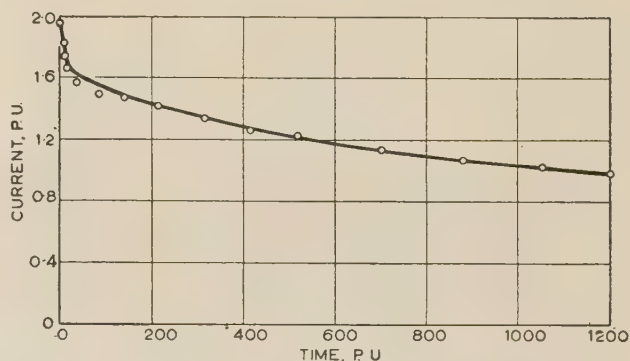
Fig. 10.—Envelope of the short-circuit current from no-load condition.
——— Theoretical. ○ Simulator.

In general, it may be said that the overall accuracy of the simulator is within $\pm 7\%$, but at a certain point, e.g. the initial value of current at the beginning of the subtransient period and when the amplitudes become small, the error with that involved in the analysis of the oscillograms may be as high as $\pm 10\%$. However, this error is relatively unimportant, since information concerning the constants of synchronous machines is not available within these limits, since it is frequently derived from oscillograph results.

(8) CONCLUSIONS

The simulator which has been described above is satisfactory for a direct representation, on an a.c. network analyser, of the alternating component of the short-circuit current of a synchronous machine. It may be used for direct solutions of multi-machine problems on a network analyser, there being no need to solve specifically for the angular relationships between the axes of the individual machines.

For a more complete simulation, other factors have to be

12

considered. Thus, a representation of the d.c. component of the short-circuit current is of importance for fault analysis and protective-gear applications; the influence of voltage-regulator action may be significant in some cases. It is considered that the association of this apparatus with electronic or electro-mechanical phase-modulating equipment for the automatic solution of power-system swing-curve equations is likely to be fruitful.

(9) ACKNOWLEDGMENTS

(10) REFERENCES

(1) CRARY, S. B.: 'Power System Stability, I and II' (Wiley, New York, 1947).
(2) DOHERTY, R. E., and NICKLE, C. A.: 'Synchronous Machines: I and II—An Extension of Blondel's Two-Reaction Theory, III—Torque-Angle Characteristics under Transient Conditions, IV—Single-Phase Short-Circuits, V—Three-Phase Short-Circuits', *Transactions of the American I.E.E.*, 1926, **45**, p. 912; 1927, **46**, p. 1; 1928, **47**, p. 457 and 1930, **49**, p. 700.
(3) PARK, R. H.: 'Two-Reaction Theory of Synchronous Machines, Part I', *ibid.*, 1929, **48**, p. 716.
(4) PARK, R. H.: 'Two-Reaction Theory of Synchronous Machines, Part II', *ibid.*, 1933, **52**, p. 352.
(5) CONCORDIA, C.: 'Synchronous Machines, Theory and Performance' (Wiley, New York, 1951).
(6) ADKINS, B.: 'Transient Theory of Synchronous Generators connected to Power Systems', *Proceedings I.E.E.*, Paper No. 1051 S, November, 1950 (**98**, Part II, p. 510).
(7) DOHERTY, R. E.: 'A Simplified Method of Analyzing Short-Circuit Problems', *Transactions of the American I.E.E.*, 1923, **42**, p. 841.
(8) CLARKE, E.: 'Circuit Analysis of A.C. Power Systems', Vols. I and II (Wiley, New York, 1943 and 1950).
(9) CAMBURN, R. D., and GROSS, E. T. B.: 'Analysis of Synchronous Machine Short Circuits', *Transactions of the American I.E.E.*, 1950, **69**, Part II, p. 671.
(10) CHING, Y. K., and ADKINS, B.: 'Transient Theory of Synchronous Generators under Unbalanced Conditions', *Proceedings I.E.E.*, Monograph No. 85 S, December, 1953 (**101**, Part IV, p. 166).
(11) KU, Y. H.: 'Transient Analysis of A.C. Machinery', *Transactions of the American I.E.E.*, 1929, **48**, p. 707.
(12) KU, Y. H.: 'Transient Analysis of Rotating Machines and Stationary Networks by Means of Rotating Reference Frames', *ibid.*, 1951, **70**, Part I, p. 943.
(13) KU, Y. H.: 'Rotating-Field Theory and General Analysis of Synchronous and Induction Machines', *Proceedings I.E.E.*, Monograph No. 54 U, December, 1952 (**99**, Part IV, p. 410).
(14) ROBERT, M. R.: 'Micromachines and Micronetworks, Study of the Problems of Transient Stability by the Use of Models Similar Electromechanically to Existing Machines and Systems', C.I.G.R.É., Paris, 1950, Paper No. 338.
(15) CONCORDIA, C.: 'The Differential Analyser as an Aid in Power-System Analysis', *ibid.*, Paper No. 311.
(16) ALDRED, A. S., and DOYLE, P. A.: 'Electronic Analogue-Computer Study of Synchronous-Machine Transient Stability', *Proceedings I.E.E.*, Paper No. 2015 S, June, 1956 (**104** A, p. 152).
(17) ADAMSON, C., and EL-SERAFI, A. M. S.: 'Representation [of] Saliency on A.C. Network Analysers', *ibid.*, Monograph No. 198 S, September, 1956 (**104** C, p. 108).
(18) KRON, G.: 'Equivalent Circuits of Electrical Machinery' (Wiley, New York, 1951).
(19) ADAMSON, C., BARNES, L., and NELLIST, B. D.: 'The Automatic Solution of Power-System Swing-Curve Equations', *Proceedings I.E.E.*, Paper No. 2035 S, May, 1956 (**104** , p. 143).

(11) APPENDICES

(11.1) The Equations of the Synchronous Machine

Eqn. (1) is the matrix equation, in per-unit terms, of the salient-pole synchronous machine connected through an external impedance to a busbar, on the assumptions of the two-reaction theory. The convention for the signs of voltage and current is that V is the impressed voltage at the terminals, and that positive current I corresponds to generation; also, the rotational voltage in the direct axis, due to the quadrature axis, is regarded as positive, whilst that in the quadrature axis, due to the direct axis, is negative for positive current.

Assuming that the speed of the machine is constant and synchronous and that the excitation is constant, then

$$p\theta = 1$$
$$G_{(p)}p = 0$$
$$G_{(p)}p\theta V_f = (X_{ad}/R_{fd})V_f = E$$

Eqn. (1) then becomes

$$
\begin{bmatrix} V_d \\ V_q - E \end{bmatrix} =
\begin{bmatrix} -(R+R_e)-p[X_e+X_{d(p)}] & [X_e+X_{q(p)}] \\ -[X_e+X_{d(p)}] & -(R+R_e)-p[X_e+X_{q(p)}] \end{bmatrix}
\begin{bmatrix} I_d \\ I_q \end{bmatrix} \quad \cdots
$$

Exact analytical solution is difficult, but a simple approximate one can be found in the literature;[5,6] this gives the current in both axes as follows:

$$
I_d = V_{q0}\left[\left(\frac{1}{X_d'' + X_e} - \frac{1}{X_d' + X_e} \right) \varepsilon^{-t/T_d''} \right.
$$
$$
\left. + \left(\frac{1}{X_d' + X_e} - \frac{1}{X_d + X_e} \right) \varepsilon^{-t/T_d'} \right]
$$
$$
+ \frac{E}{X_d + X_e} - \frac{V_{q0}}{X_d'' + X_e}\varepsilon^{-t/T_a}\cos t + \frac{V_{d0}}{X_d'' + X_e}\varepsilon^{-t/T_a}\sin t .
$$

$$
I_q = -V_{d0}\left[\left(\frac{1}{X_q'' + X_e} - \frac{1}{X_q + X_e} \right) \varepsilon^{-t/T_q''} \right]
$$
$$
+ \frac{V_{d0}}{X_q'' + X_e}\varepsilon^{-t/T_a}\cos t + \frac{V_{q0}}{X_q'' + X_e}\varepsilon^{-t/T_a}\sin t
$$

The resultant short-circuit current in phase 'a' is thus:
$$
I_a = I_d \cos(t + \theta_0) - I_q \sin(t + \theta_0)
$$

$$
I_a = \left[\left(\frac{1}{X_d'' + X_e} - \frac{1}{X_d' + X_e} \right) \varepsilon^{-t/T_d''} \right.
$$
$$
\left. + \left(\frac{1}{X_d' + X_e} - \frac{1}{X_d + X_e} \right) \varepsilon^{-t/T_d'} \right] V_{q0}\cos(t + \theta_0)
$$
$$
+ \frac{E}{X_d + X_e}\cos(t + \theta_0) - \frac{V_{q0}}{X_m}\varepsilon^{-t/T_a}\cos\theta_0
$$

$$-\frac{V_{q0}}{X_n}\varepsilon^{-t/T_a}\cos(2t+\theta_0)$$

$$+\left(\frac{1}{X_q''+X_e}-\frac{1}{X_q+X_e}\right)\varepsilon^{-t/T_q''}V_{d0}\sin(t+\theta_0)$$

$$-\frac{V_{d0}}{X_m}\varepsilon^{-t/T_a}\sin\theta_0+\frac{V_{d0}}{X_n}\varepsilon^{-t/T_a}\sin(2t+\theta_0)\qquad.\quad(5)$$

In order to solve for the fundamental-frequency component only, a great degree of simplification can be achieved by omitting the terms,

$$-\left\{(R+R_e)+p[X_e+X_{d(p)}]\right\}\text{ and }$$
$$-\left\{(R+R_e)+p[X_e+X_{q(p)}]\right\}$$

eqn. (2) then becomes

$$\begin{bmatrix}V_d\\V_q-E\end{bmatrix}=\begin{bmatrix}0 & [X_e+X_{q(p)}]\\-[X_e+X_{d(p)}] & 0\end{bmatrix}\begin{bmatrix}I_d\\I_q\end{bmatrix}\quad(6)$$

and the solution of eqn. (6) is

$$I_d=V_{q0}\left[\left(\frac{1}{X_d''+X_e}-\frac{1}{X_d'+X_e}\right)\varepsilon^{-t/T_d''}\right.$$

$$\left.+\left(\frac{1}{X_d'+X_e}-\frac{1}{X_d+X_e}\right)\varepsilon^{-t/T_d'}\right]+\frac{E}{X_d+X_e}\quad(7)$$

$$I_q=-V_{d0}\left[\left(\frac{1}{X_q''+X_e}-\frac{1}{X_q+X_e}\right)\varepsilon^{-t/T_q''}\right]\qquad.\quad(8)$$

It will be observed that the fundamental-frequency terms in eqns. (3) and (4) are identical with the solutions for I_d and I_q in eqns. (7) and (8). It thus follows that eqn. (6) forms a simple and satisfactory basis for the simulator.

1.2) Establishment of the Equivalent Circuits of the Synchronous Machine

By writing down the equations of the equivalent circuits shown in Fig. 3B:

If, in eqn. (6), the resistance term $R+R_e$ is introduced, then, after re-arrangement, this equation becomes:

$$\begin{bmatrix}V_d+(R+R_e)I_d-X_eI_q\\V_q-E+(R+R_e)I_q+X_eI_d\end{bmatrix}=\begin{bmatrix}0 & X_{q(p)}\\-X_{d(p)} & 0\end{bmatrix}\begin{bmatrix}I_d\\I_q\end{bmatrix}(11)$$

By combining the busbar voltage V, the voltage drop in the resistance, and the voltage drop in the external impedance, the voltage behind the resistance of the machine, V_R, is obtained. V_R has the components V_{Rd} and V_{Rq}, and the equation of the machine is now:

$$\begin{bmatrix}V_{Rd}\\V_{Rq}-E\end{bmatrix}=\begin{bmatrix}0 & X_{q(p)}\\-X_{d(p)} & 0\end{bmatrix}\begin{bmatrix}I_d\\I_q\end{bmatrix}\quad.\quad(12)$$

Eqn. (12) is exactly the same as the matrix equation of the equivalent circuits, if the components of the voltage behind the resistance of the machine, V_{Rd} and V_{Rq}, are applied to the terminals in addition to the signal E. Thus, such equivalent circuits are suitable for the solution of the fundamental-frequency component of the transient, the resistance of the machine and the external impedance being taken into account.

It should be noted that the reintroduction of resistance, leading to eqn. (11) above, has no effect on the principle of the simulator and is independent of the fact that the terms

$$-\left\{(R+R_e)+p[X_e+X_{d(p)}]\right\}\text{ and }$$
$$-\left\{(R+R_e)+p[X_e+X_{q(p)}]\right\}$$

were neglected in order to extract the fundamental-frequency solution only of eqn. (2). The asymmetrical d.c. and second-harmonic terms are a consequence not of resistance, but of interaction between direct and quadrature axes due to rotational voltages.[6] If the resistance terms in the two expressions above had alone been retained in eqn. (2) then the solution would still have been for fundamental-frequency terms only;[5] however, in this case the magnitude of the current is reduced and its phase slightly altered, since the current is encountering the total impedance and not merely the reactance of the machine.

$$V_{Rd}=I_q\left[X_q-X_{aq}+\frac{X_{aq}(X_{kq}-X_{aq}+R_{kq}/p)}{X_{kq}+R_{kq}/p}\right]=I_q\left(X_q-\frac{X_{aq}^2p}{X_{kq}p+R_{kq}}\right)=I_qX_{q(p)}\quad.\quad.\quad.\quad.\quad.\quad(9)$$

$$V_{Rq}-E=\left[(X_d-X_{ad}+\frac{X_{ad}(X_{kd}-X_{ad}+R_{kd}/p)(X_{fd}-X_{ad}+R_{fd}/p)}{(X_{kd}-X_{ad}+R_{kd}/p)X_{ad}+(X_{fd}-X_{ad}+R_{fd}/p)X_{ad}+(X_{kd}-X_{ad}+R_{kd}/p)(X_{fd}-X_{ad}+R_{fd}/p)}\right]I_d$$

$$=I_d\left[X_d-\frac{(X_{fd}p+R_{fd})X_{ad}^2p+(X_{kd}p+R_{kd})X_{ad}^2p-2X_{ad}^3p^2}{(X_{kd}p+R_{kd})(X_{fd}p+R_{fd})-X_{ad}^2p^2}\right]=I_dX_{d(p)}\quad.\quad(10)$$

AN EXPERIMENTAL INVESTIGATION OF TOOTH-RIPPLE FLUX PULSATIONS II SMOOTH LAMINATED POLE-SHOES

By Prof. J. GREIG, M.Sc., Ph.D., Member, and K. C. MUKHERJI, B.E., Ph.D., Graduate.

SUMMARY

The paper describes certain experiments designed to investigate the mode of penetration of tooth-ripple flux pulsations in smooth laminated pole-shoes. Simple measurements made by means of search coils embedded in holes drilled axially at different radial depths from the pole face have revealed that a dynamic mode of flux pulsation, which corresponds to the static tooth-flux distribution, persists, at significant amplitude, at the normal operating speed of the machine. Measurements have been made with two sets of laminations of different thickness and material at a range of tooth-ripple frequencies. The results provide substantial evidence in support of an analysis of tooth-ripple phenomena contained in a companion paper.

(1) LIST OF PRINCIPAL SYMBOLS

The rationalized M.K.S. system of units is used, unless otherwise stated.

d = Classical depth of penetration.
f_t = Tooth-ripple frequency, c/s.
h = Lamination thickness.
$k = 2\pi/\lambda$.
λ = Armature slot-pitch.
μ = Incremental permeability.
ρ = Resistivity.
ω = Angular tooth-ripple frequency.

The symbols β_0 and R_z are explained in the text.

(2) INTRODUCTION

The initial experiments which resulted in the series of measurements reported in the paper and in the theoretical investigation of the companion paper[7] by Prof. H. Bondi and K. C. Mukherji, were prompted by a consideration of the theoretical work on tooth-ripple losses by Dr. W. J. Gibbs.[4] Gibbs's theoretical treatment for laminated pole-shoes assumed that tooth-ripple flux pulsations would suffer attenuation on passing in from the pole face at a rate corresponding to the classical depth of penetration.* Consideration of the physics of the problem raised doubts as to the validity of this assumption, and it was decided to make a simple experiment on a machine which had been built for work on tooth ripples. This experiment, which has already been partially reported,[6] consisted in the measurement of tooth-ripple-frequency electromotive forces induced in search coils embedded in axial holes in a laminated pole-piece at a range of depths behind the pole face. The classical depth of penetration

* The 'depth of penetration' is here defined as $1/\sqrt{2}$ times the depth from the pole face in which the amplitude of the tooth-ripple flux pulsations attenuates to $1/\varepsilon$ times the value at the pole face.

for the conditions of the experiment was less than a millimetr but an e.m.f. of readily measurable magnitude was found to l induced in search coils at depths greater than ten times tl classical depth of penetration. Following this observation, more extensive series of experiments was initiated. A fre attack was made on the theoretical problem and the existii literature was reviewed.

It was found that all theoretical treatments were in agreeme about the rate of radial attenuation of the tooth-ripple fl pulsations in the idealized case of an infinitely long solid pol shoe, namely that it was dependent only on the permeabili and resistivity of the material of the pole shoe, and on the toot ripple frequency $f_t (= \omega/2\pi)$; i.e. the depth of penetration is tl classical one, d, where

$$d = \left(\frac{\rho}{\mu\omega}\right)^{1/2}$$

Similar unanimity was not found, however, in the publish literature dealing with the problem of the laminated pole-shc A 'skin-effect' phenomenon along the thickness of each lamin tion in a laminated pole-shoe had been anticipated by Adar and his associates.[1] Spooner and Kinnard[3] had held a simil view; they had also suggested that, for laminated materials su as were commonly used in commercial machines, 'the depth penetration of the pulsating flux is only slightly affected by tl tooth-ripple frequency and is determined almost completely the dimensions of the slots, air-gap, etc'.* F. W. Carter[2] ai Gibbs,[4] on the other hand, had assumed the classical depth penetration to apply to the case of a laminated pole-shoe well. Prof. G. W. Carter's[5] treatment took no account of ai skin effect along the thickness of a lamination. However, it w one of the conclusions of his analysis that the depth of penetratio of the tooth-ripple field in a very finely laminated pole-shoe w not the classical one; it was determined by the lamination thic ness rather than the electromagnetic properties of the material the tooth-ripple frequency.

(3) THE EXPERIMENTAL MACHINE

The machine employed in the investigation had six salie poles with suitable field coils, and an unwound rotor with semi-closed slots (see Fig. 1). The design figure for the me air-gap induction for an output of 30 kVA at 50 c/s, correspondi to a speed of 1 000 r.p.m., was 0·68 Wb/m². Two sets of pc punchings were available, namely 3·18 mm-thick cold-roll close-annealed sheet steel and 0·40 mm-thick Losil 19. rectangular strip was cut out of each set for the measureme of the resistivity of the material. Ring specimens were us for obtaining minor loops with a polarization of 0·68 Wb/m² an automatic d.c. hysteresis-loop recorder. The slope of t major axis of a minor loop was taken as the incremental pe meability for calculating the classical depth of penetration different frequencies. These results for the two sets of punchin are given in Table 1.

* See Reference 3, p. 265.
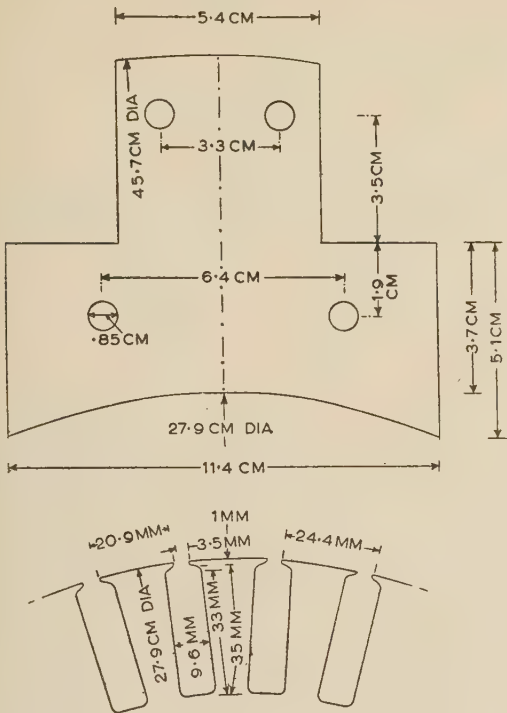
Fig. 1.—Pole punching and rotor stamping.

Length of pole arc: 11·78 cm.
Rotor material: 0·40 mm Losil 19.

for the search coils. It was recognized that the finite size of the holes might have an appreciable effect on measurements, and an attempt was made to repeat certain series of measurements with the search wires inserted into finer holes. It has, so far, been practicable to assemble experimental pole-pieces with ten aligned holes of $\frac{1}{32}$ in diameter.

(5) PRELIMINARY EXPERIMENTS

The first set of pole-pieces was built with the 3·18 mm-thick punchings, one of these six being used as the experimental pole-piece. It had three search coils, B, C and D (see Fig. 2), threaded through three pairs of holes, the centres of which were at successive depths of 1·5, 3·0 and 5·0 mm from the pole face. The pairs of holes at the different depths were staggered along the pole arc in order to avoid disturbing the flux paths more than necessary. A fourth coil A, half a slot pitch wide, was glued to the pole face at the middle of the pole arc, while a fifth coil G was wound round the tips of the pole (see Fig. 2). The latter was used for measuring the total air-gap flux under the experimental pole by means of a fluxmeter, and also for measuring the e.m.f. due to the reluctance pulsation of this main flux over the entire cross-section of the pole.

The waveform of the e.m.f. induced in each search coil was observed on a cathode-ray oscilloscope, and the r.m.s. value of the fundamental component was measured by means of an electronic wave analyser. Measurements[6] were made with a constant mean air-gap induction, B_m, of 0·64 Wb/m² at three different tooth-ripple frequencies, namely 90, 600 and 900 c/s, corresponding, respectively, to speeds of 150, 1 000 and 1 500 r.p.m. of the experimental machine.

Table 1

CLASSICAL DEPTH OF PENETRATION $d = \left(\dfrac{\rho}{\mu\omega}\right)^{1/2}$

Material	Resistivity	Incremental permeability	Angular frequency	Depth of penetration
	ohm-m	H/m	rad/sec	mm
Cold-rolled close-annealed sheet steel	$11·6 \times 10^{-8}$	$360 \times 4\pi \times 10^{-7}$	$90 \times 2\pi$ $600 \times 2\pi$ $900 \times 2\pi$	0·68 0·26 0·21
Losil 19..	$35·3 \times 10^{-8}$	$800 \times 4\pi \times 10^{-7}$	$90 \times 2\pi$ $600 \times 2\pi$ $900 \times 2\pi$	0·79 0·31 0·25

(4) EXPERIMENTAL TECHNIQUE

It was thought that, if the depth of penetration of tooth-ripple flux pulsation in a laminated pole-shoe should be substantially larger than the classical one, it might be possible to investigate the attenuation of the ripples with radial depth by locating search coils in holes drilled in a laminated pole-shoe at different depths behind the pole face. Single-turn search coils having a peripheral width of half a slot pitch (12·2 mm) and running the full axial length of the pole were used. This was considered to be preferable to the use of single search wires with a return path outside the machine, as the flux linkage with single search wires would be less definite and might be complicated by main-flux reluctance pulsation phenomena. The e.m.f. induced in a half-slot-pitch-wide search coil would comprise the fundamental and all other odd harmonic components; the even harmonic components would, however, be absent.

Although the above technique was simple in principle, the mechanical task of assembling a pole-piece proved to be difficult in practice. In the first instance, $\frac{1}{16}$ in-diameter holes were used
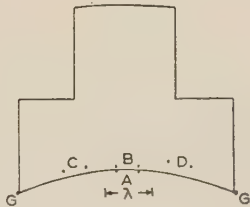
Fig. 2.—Arrangement of search coils in preliminary experiment.

Depth of coil B = 1·5 mm.
Depth of coil C = 3·0 mm.
Depth of coil D = 5·0 mm.

(6) PRELIMINARY RESULTS AND COMMENTS

The measurements (see Table 2 and Fig. 3) demonstrated the existence of substantial tooth-ripple flux pulsation at a depth as great as 5 mm from the pole face. That these were not primarily main-flux reluctance pulsations was shown by the relatively small

Table 2

RESULTS OF PRELIMINARY DEPTH-OF-PENETRATION MEASUREMENTS

Uninsulated $3 \cdot 18$ mm thick cold-rolled close-annealed punchings. $B_m = 0 \cdot 64$ Wb/m^2.

Search coil	Depth from pole face	Fundamental induced e.m.f.					
		90 c/s		600 c/s		900 c/s	
	mm	mV (r.m.s.)	%	mV (r.m.s.)	%	mV (r.m.s.)	%
A	0	47·5	100·0	320·0	100·0	470·0	100·0
B	1·5	24·0	50·5	175·0	54·6	255·0	54·2
C	3·0	15·0	31·6	111·0	34·7	170·0	36·2
D	5·0	11·0	23·2	72·5	22·6	113·0	24·0
G	Round pole tips	5·5	11·6	42·5	13·1	62·0	13·2
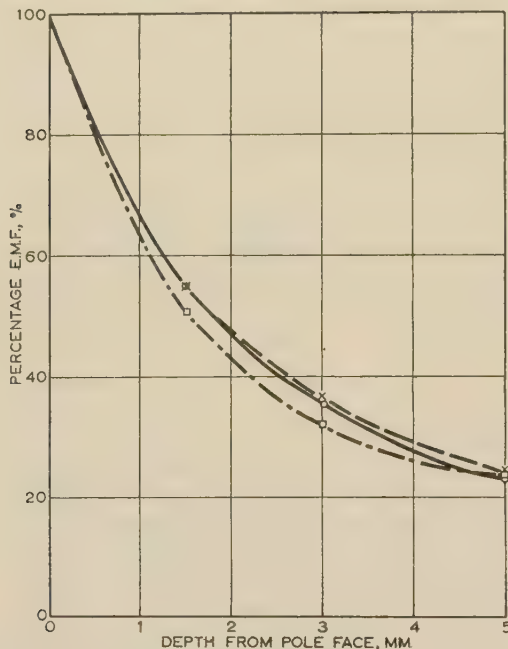
Fig. 3.—Preliminary results of depth-of-penetration measurements.

$3 \cdot 18$ mm uninsulated cold-rolled close-annealed pole punchings.

90 c/s ---□---□--- 150 r.p.m.
600 c/s ——○——○—— 1 000 r.p.m.
900 c/s --×--×-- 1 500 r.p.m.

induced e.m.f. in the search coil G embracing the whole pole. This e.m.f. at the pole face, reduced in proportion to the areas embraced by the search coils A and G [i.e. in the ratio (half slot pitch)/(pole arc), or 12·2 : 117·8] represented at 600 c/s only about 1·4% of the e.m.f. induced in the search coil A. Thus more than 20% of the tooth-ripple at the pole face penetrated as deep as 5·0 mm into the laminated pole at a frequency of 600 c/s. The corresponding classical depth of penetration for the material employed had been calculated to be only 0·26 mm. Similar deep penetration was observed at tooth-ripple frequencies as widely separated as 90 and 900 c/s (see Fig. 3), for which the classical penetration depths were 0·68 and 0·21 mm, respectively. Still more striking was the observation that the measured values of the fundamental components of all search-coil e.m.f.'s were approximately proportional to the tooth-ripple frequency for a given mean air-gap induction; i.e. the rate of radial attenuation of the fundamental tooth-ripple flux pulsation remained substantially unaltered over the wide range of frequency (90–900 c/s) for which

measurements had been made. This was incompatible with the classical concept, according to which the attenuation rate should have varied directly as the square root of the tooth-ripple frequency.

Doubling a small exciting current in the field coils of the experimental machine resulted in an approximate doubling of the search-coil e.m.f.'s. It was hardly likely, therefore, that the observed phenomena arose as any second-order effect of non-linear electromagnetic properties of the material of the pole punchings. There remained, of course, an element of uncertainty regarding the quantitative effect of the holes that accommodated the search coils.

These observations led to a search for alternative solutions of Maxwell's field equations which would be physically tenable along with the classical solution for the tooth-ripple field. The ultimate result has been a detailed quantitative analysis of the entire tooth-ripple problem, which forms the subject of Reference 7.

(7) THEORETICAL PREDICTION OF EXPERIMENTAL RESULTS

In general terms, the conclusions of the theoretical analysis can be summed up as follows:

(a) Laminating a pole shoe effectively results in a fraction of the total tooth-ripple field at the pole face penetrating much deeper radially within the pole shoe than the classical depth of penetration.

(b) The thinner the laminations, the larger is this fraction.

(c) The mode of deep radial penetration is the same as the static mode and depends solely on the armature slot-pitch.

(d) The deep radial penetration is accompanied by a 'skin effect' along the thickness of each lamination.

It is possible to predict, on the basis of the analysis, some results of practicable measurements concerning the mode of deep radial penetration of the tooth-ripple field. If h represents the lamination thickness and $k = 2\pi/\lambda$, where λ is the armature slot-pitch, the above-mentioned 'fraction' can be shown to be approximately equal to $1/\beta_0$, where

$$\beta_0 = -j + \frac{2\sqrt{2}}{\pi^2} kh[(1 - \sqrt{2}kd) + j] \coth\left[(1 + j)\frac{h}{2\sqrt{2}d}\right]$$

$$\sum_{n=0}^{\infty} \frac{\left[j + (2n+1)^2 \frac{\pi^2 d^2}{h^2}\right]^{-3/2}}{(2n+1)^2}$$

$$+ (1 - j)\frac{4\sqrt{2}}{\pi^4} k^2 h^2 \sum_{n=0}^{\infty} \frac{\left[j + (2n+1)^2 \frac{\pi^2 d^2}{h^2}\right]^{-1}}{(2n+1)^4}$$

The radial attenuation of this part of the field follows the exponential law ε^{-kz}, z being the radial depth from the pole face. Thus the ratio R_z between the r.m.s. value of the

.n.f. induced in a half-slot-pitch-wide search coil embedded
.a radial depth z ($z \gg d$) from the pole face and that of the
.n.f. induced in a similar coil on the pole face is obtained as

$$R_z = \frac{100}{|\beta_0|}\varepsilon^{-kz} \text{ per cent}$$

.ere $|\beta_0|$ indicates the modulus of β_0. Tables 3 and 4 show the
.culated values of $1/|\beta_0|$ and of R_z. These values of R_z are also
.tted in Fig. 4.

The experimental results relate to the total tooth-ripple field
at any given depth from the pole face, while R_z relates to the
mode of deep radial penetration alone. It is thus only at
sufficiently large depths from the pole face, i.e. for $z \gg d$, that
experimental results could approximate closely to the calculated
figures.

(8) FURTHER EXPERIMENTS

Once theory in general agreement with the preliminary experi-
mental observations had been established by means of the

Table 3

CALCULATED VALUES OF $\dfrac{1}{|\beta_0|}$

$\lambda = 2 \cdot 44$ cm. $k = 258$ m^{-1}

Material	h	ω	d	$\dfrac{1}{\beta_0}$	$\dfrac{1}{|\beta_0|}$
	mm	rad/s	mm		
Cold-rolled close-annealed sheet steel	3·18	$90 \times 2\pi$ $600 \times 2\pi$ $900 \times 2\pi$	0·68 0·26 0·21	$0 \cdot 118 + j0 \cdot 802$ $0 \cdot 038 + j0 \cdot 720$ $0 \cdot 031 + j0 \cdot 714$	0·813 0·721 0·714
Losil 19	0·40	$90 \times 2\pi$ $600 \times 2\pi$ $900 \times 2\pi$	0·79 0·31 0·25	$0 \cdot 0 + j1 \cdot 0$ $0 \cdot 0 + j1 \cdot 0$ $0 \cdot 0 + j1 \cdot 0$	1·0 1·0 1·0

Fig. 4.—Calculated deep penetration of tooth-ripple field.

——— 0·40 mm Losil 19.
—·—·— 3·18 mm cold-rolled close-annealed sheet steel, 90 c/s.
— — — 3·18 mm cold-rolled close-annealed sheet steel, 600 and 900 c/s.
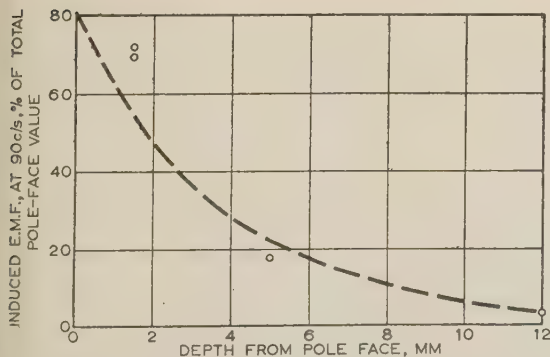
analysis, it remained to confirm this and to extend the range of
the observations. The principle of the measurements was as in
the preliminary experiment, the same two sets of pole punchings

Table 4

CALCULATED VALUES OF R_z

$\lambda = 2 \cdot 44$ cm $k = 258$ m^{-1}

Depth from pole face, z	$R_z = \dfrac{100}{|\beta_0|}\varepsilon^{-kz}$			
	3·18 mm cold-rolled close-annealed sheet steel			0·40 mm Losil 19
	90 c/s	600 c/s	900 c/s	90, 600 and 900 c/s
mm	%	%	%	%
0	81·30	72·05	71·42	100·0
1·5	55·30	49·00	48·50	68·00
3·0	37·50	33·20	32·90	46·10
5·0	22·40	19·89	19·65	27·60
6·0	17·30	15·38	15·19	21·20
8·0	10·29	9·13	9·00	12·63
10·0	6·16	5·46	5·40	7·58
12·0	3·68	3·26	3·23	4·52

being used, but with the thicker stampings an interlaminar
insulation of thin tissue paper was introduced. An endeavour
was also made to determine whether the size of the holes that
accommodated the search coils had any significant effect on the
measurements.

(8.1) Experiments with 3·18 mm Cold-Rolled Close-Annealed Punchings

The measurements of search-coil induced e.m.f., as in the
preliminary experiment, were repeated with a new pole-piece built
with the 3·18 mm-thick pole punchings, first without and next
with interlaminar paper insulation. The $\frac{1}{16}$ in-diameter holes for
the search coils were drilled directly behind one another at depths
of 1·5, 5, and 12 mm, symmetrically about the centre-line of the
pole. The experimental results are shown in Figs. 5–8. Search

Fig. 5.—Depth-of-penetration measurement.

3·18 mm insulated cold-rolled close-annealed pole punchings, 90 c/s.
O Experimental result.
— — — Calculated result for 'deep' mode of penetration only, R_z.
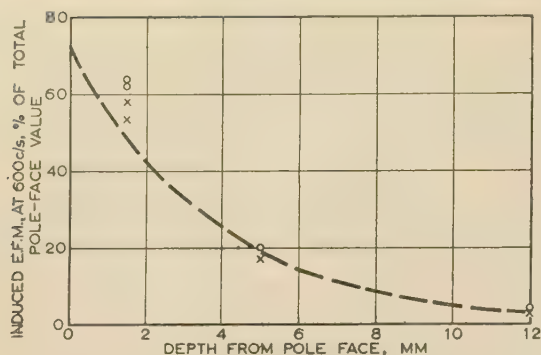
Fig. 6.—Depth-of-penetration measurement.

3·18 mm cold-rolled close-annealed pole punchings, 600 c/s.
○ Experimental result for insulated punchings.
× Experimental result for uninsulated punchings.
- - - - Calculated result for 'deep' mode of penetration only, R_z.
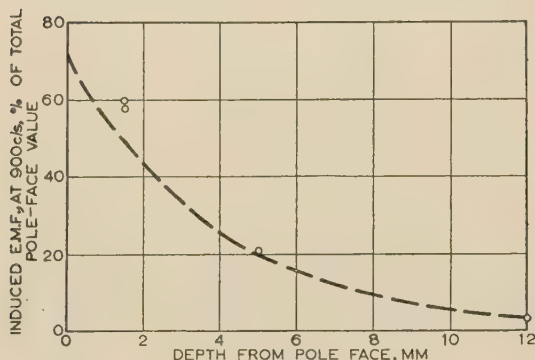
Fig. 7.—Depth-of-penetration measurement.

3·18 mm insulated pole punchings, 900 c/s.
○ Experimental result.
- - - - Calculated result for 'deep' mode of penetration only, R_z.
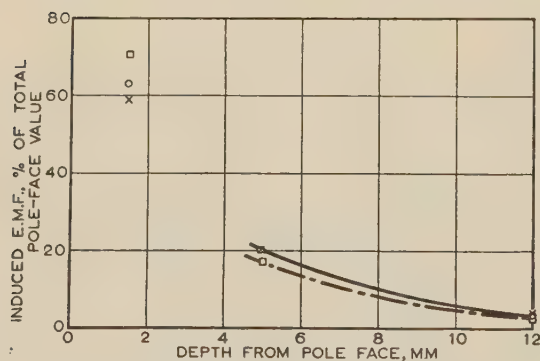
Fig. 8.—Penetration at different frequencies.

3·18 mm insulated cold-rolled close-annealed pole punchings.
□ 90 c/s.
○ 600 c/s.
× 900 c/s.

coils in holes offset along the pole arc were utilized in setting the air-gap, which was adjusted until the e.m.f.'s induced in these coils did not differ by more than 10%. Indeed, this pro-cedure for obtaining a nearly uniform gap was essential for the measurements, since the ratios of the e.m.f.'s induced in the various search coils were found to be extremely sensitive to non-uniformity of the air-gap.

(8.2) Experiments with 0·40 mm Losil Punchings

Another experimental pole-piece was built with 0·40 mm-thick Losil 19 pole punchings, the $\frac{1}{16}$ in-diameter holes for the

search coils being drilled behind one another. The experimen[tal] results with coils at depths of 1·5, 3 and 10 mm symmetrical ab[out] the centre line of the pole are shown in Fig. 9. In view of the la[rge]
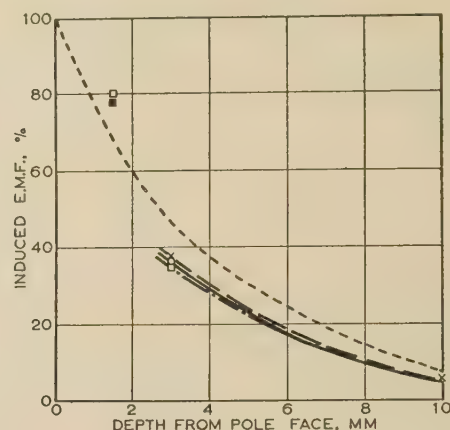
Fig. 9.—Penetration at different frequencies in Losil 19 pole punchings.

—·—□—·— 90 c/s } Experimental
———○——— 600 c/s } results.
- - -×- - - 900 c/s }
············ Calculated R_z for 90, 600 and 900 c/s.

number (over 400) of lamination interfaces, no special inte[r]-laminar insulation was used; also, it was assumed that the fa[ct] that the five other pole-pieces of the experimental machine we[re] built with thick punchings had a negligible effect on measur[e]-ments of tooth-ripple phenomena in the one odd pole-pie[ce] constructed with thin Losil punchings.

(8.3) Experiments concerning the Effect of Holes

It was considered desirable to try to ascertain whether th[e] e.m.f.'s induced in the search coils were appreciably affected [by] the sizes of the holes in which the coil sides were located. F[or] this purpose, the measurements were repeated first with $\frac{3}{32}$ i[n-] diameter holes in a new experimental pole-piece of 3·18 m[m] thick punchings and finally with another similar pole in whi[ch] $\frac{1}{32}$ in-diameter holes had been drilled.

It will be seen from the results illustrated in Figs. 10 and 1[1] that reduction of hole diameter from $\frac{3}{32}$ to $\frac{1}{16}$ in diminishe[d] slightly the deviation from the theoretical curve, but that furth[er] reduction to $\frac{1}{32}$ in diameter produced no appreciable change.

(8.4) Tests with Search Coils in a Radial Plane

Some observations relating to the radial attenuation of th[e] tangential component of the tooth-ripple flux density were ma[de] by forming search coils in a radial plane, utilizing existing holes [at] depths of 3, 6 and 10 mm, in one of the cold-rolled close-anneal[ed] sheet steel pole-pieces. These coils linked with the tooth-ripp[le] flux between the pairs of wires at depths 0 and 3 mm, 3 and 6 m[m] and 6 and 10 mm, respectively, from the pole face. Th[e] measured fundamental components of the e.m.f.'s induced i[n] these coils were found to be in the ratio 58·3 : 26·7 : 15·[0] The results of the harmonic analysis of the e.m.f. induced in [a] radial coil of depth 10 mm (between depths of 0 and 10 mm fro[m] the pole face) are compared in Table 5 with those for the e.m.[f.] of a half-slot-pitch-wide coil on the pole face.

The even-harmonic residues in the half-slot-pitch-wide peri[-] pheral coil are presumably due to slight inaccuracies of geometr[y.] It will be noticed that the theoretical ratio of 2 : 1 for the funda[-]
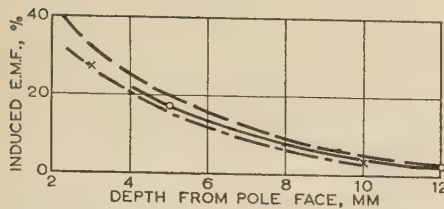
Fig. 10.—Depth-of-penetration measurements with different sizes of holes.

3·18 uninsulated cold-rolled close-annealed pole punchings at 600 c/s.

————O———— $\frac{1}{16}$ in holes ⎱ Experimental
—·—×—·— $\frac{3}{32}$ in holes ⎰ results.
——————— Calculated R_z.

Fig. 11.—Depth-of-penetration measurement.

3·18 mm insulated cold-rolled close-annealed pole punchings, 600 c/s.

————O———— $\frac{1}{16}$ in holes ⎱ Experimental
×⠀⠀⠀⠀⠀$\frac{1}{16}$ in holes ⎰ results.
——————— Calculated result for 'deep' mode of penetration only, R_z.

...ental and odd-harmonic components of the e.m.f. induced in ...ch a peripheral coil to the e.m.f. which would be induced in ...radial coil embracing the whole flux variation is approximated ... the experimental results.

Table 5

	Induced e.m.f.				
	Fundamental (600 c/s)	2nd harmonic	3rd harmonic	4th harmonic	5th harmonic
Peripheral pole-face coil (width $= \frac{\lambda}{2} = 12\cdot2$ mm)	mV (r.m.s.) 210	mV (r.m.s.) 14	mV (r.m.s.) 150	mV (r.m.s.) 6	mV (r.m.s.) 96
Radial coil (between depths 0 and 10 mm)	115	100	73	72	50

(8.5) Test for Reluctance Pulsation

A further test was made to determine the uniformity of the ...lsations of the main flux round the magnetic circuit. It ...as found, however, that the amplitude of the flux pulsations ...nking with a search coil at the pole face embracing the whole ...ole was reduced to a quarter of its value when the coil was ...oved back 1 cm from the pole face. It was concluded that ...large part of the e.m.f. induced in this coil, when at the pole ...ce, resulted from a 'flux swing', and that it was hardly justifi-...ole to make allowance for the relatively small proportion of the ...m.f. induced in the half-slot-pitch-wide search coils resulting ...om reluctance pulsation.

(9) DISCUSSION OF RESULTS AND CONCLUSION

It has not been possible to demonstrate theoretically the effect ... the sizes of the holes accommodating the search coils in the ...ove series of measurements. It is, however, obvious from ...nysical considerations that the holes would have two effects ...artially counteracting each other; the increased reluctance of ...e flux paths would tend to reduce the flux linkage of a search ...oil, while the removal of a sheet of eddy currents would help ... increase this linkage. The absence of substantial variation in ...sults for the range of hole size between $\frac{1}{2}$ and $\frac{3}{32}$ in suggests ...at the effect is small, although this is in no sense conclusive. ...he effect of introducing interlaminar insulation with the ...old-rolled close-annealed sheet-steel punchings is also seen to ... not very marked, although it did lead to results more in line ...ith analytical prediction. Measurements at 90 c/s were, in ...neral, liable to errors larger than those at the higher frequencies ...cause of the rather small e.m.f.'s available at that low fre-

quency. However, there is general agreement between the attenuation observed experimentally and the theoretically predicted curves.

In comparing the experimental observations with theoretically predicted values, it must be remembered that the theoretical analysis applies to an ideal cylindrical structure without the discontinuities produced by salient poles. In addition, the degree of uniformity which could be achieved in the air-gap was not as high as could be desired nor was the accuracy of alignment of the experimental pole-pieces themselves. Considering the actual conditions obtaining in the experiments, the general correspondence of these experimental results with theoretical predictions is believed to provide substantial evidence in support of the particular parts of the analysis outlined in Reference 7 which concern the deep-penetration phenomenon of tooth ripples in smooth laminated pole-shoes.

(11) REFERENCES

(1) ADAMS, C. A., LANIER, A. C., POPE, C. C., and SCHOOLEY, C. O.: 'Pole-Face Losses', *Transactions of the American I.E.E.*, 1909, **28**, p. 1133.

(2) CARTER, F. W.: 'Pole-face Losses', *Journal I.E.E.*, 1916, **54**, p. 168.

(3) SPOONER, T., and KINNARD, I. F.: 'Surface Iron Losses with Reference to Laminated Materials', *Transactions of the American I.E.E.*, 1924, **43**, p. 262.

(4) GIBBS, W. J.: 'Tooth-ripple Losses in Unwound Pole-shoe, *Journal I.E.E.*, 1947, **94,** Part II, p. 2.

(5) CARTER, G. W.: 'A Note on the Surface Loss in a Laminate Pole-face', *Proceedings I.E.E.*, Monograph No. 123, March 1955 (**102** C, p. 217).

(6) MUKHERJI, K. C.: Discussion on Reference 5, *ibid.*, 1956 **103** C, p. 230.

(7) BONDI, H., and MUKHERJI, K. C.: 'An Analysis of Tooth Ripple Phenomena in Smooth Laminated Pole-shoes' (at page 349).

PROPAGATION OF TRANSIENTS IN WAVEGUIDES

By A. E. KARBOWIAK, Ph.D.

SUMMARY

The basic theory of transient propagation in waveguides is formulated and developed. The theory is subsequently applied to the propagation of a unit-step-modulated carrier and the propagation of pulses. Formulae and graphs are outlined for various transients involved and pulse distortion is studied in detail. The results are illustrated by numerical examples.

LIST OF SYMBOLS

p = Complex frequency.

t = Time.

$\gamma(p) = \alpha(p) + j\beta(p)$ = Propagation coefficient of the waveguide (function of p).

$\alpha(p)$ = Attenuation coefficient.

$\beta(p)$ = Phase-change coefficient.

z = Co-ordinate in the direction of propagation.

ω = Real angular frequency.

ω_0 = Carrier angular frequency.

ω_c = Cut-off frequency of a perfect waveguide.

ω_m = Cut-off frequency of an imperfect waveguide.

u, v, s = Complex variables.

b = A coefficient defined by eqn. (43).

$V = \dfrac{z/t}{c}$ = A parameter.

A = A parameter defined by eqn. (49).

v_g = Group velocity.

$t_1 = z/v_g$ = A time parameter.

$\Phi(t)$, $f(p)$, $S(z, t)$, $P(u)$, $\Lambda(t, z)$, $\Psi(t, z, \Delta t)$, $\Omega(At, A\Delta t, A\Delta t)$
= Various functions defined in appropriate places.

l = Length of the waveguide.

(1) INTRODUCTION

During the last two decades, waveguide theory and practice have undergone a tremendous development; but the main impetus was derived from the work on radar during the last World War. It is, therefore, at first somewhat surprising that, although radar is essentially concerned with transient phenomena, no theory of transient behaviour of waveguides has as yet been formulated.* The probable explanation is that no material transient distortion that could be ascribed to the waveguide was actually observed, and that the existing theories of waveguide propagation based on c.w. propagation seemed to be adequate, while the analytically difficult task of investigating transient behaviour seemed unrewarding. Yet it is well known that a waveguide is dispersive—more so than a coaxial line—and therefore transient distortion may be tentatively anticipated. It is, however, not easy to make a judicious guess at the nature of the distortion, although we might surmise (somewhat incorrectly) that the envelope distortion would be like that of a low-pass network.

---

* Recently, however, Gajewski[11] succeeded in obtaining an expression for the precursor (corresponding to the B-precursor discussed below) of a unit-step-modulated carrier propagated in a perfect waveguide, and he included some discussion of the nature of this precursor.

The reason why most existing waveguide equipment seems to pass transients undistorted is that the waveguide lengths involved are too short and the transient durations too long for any distortion to be noticeable. Recently, however, observable distortion of pulses transmitted through a length of waveguide have been reported.[1,2]

Nevertheless, with the advent of millimicrosecond-pulse techniques[1] and possibilities of waveguide trunk communication,[3] the pulse distortion by a waveguide becomes one of its principal features; a feature which sets an upper limit to repeater spacings and becomes an important parameter governing the design of numerous components.

The aim of the paper is the formulation of the basic theory of transient propagation in waveguides under various conditions, and the determination of its application to pulse transmission.

There is a somewhat allied problem which was investigated some time ago by Sommerfeld and Brillouin,* namely the propagation of *plane* waves in unbounded homogeneous dispersive *media*. Our problem, on the other hand, is that of *modal* (E or H) propagation with dispersion due to proximity of physical *boundaries*.

The two problems are physically entirely distinct and our analytical approach is also different, but qualitatively the results have a little in common in that the signal is preceded and followed by anticipatory and posterior transients, respectively.

(2) SOME FUNDAMENTAL CONCEPTS

(2.1) Introductory

Before developing a rigorous theory of transient propagation, the problems must first be examined.

Imagine an indefinitely long waveguide line (Fig. 1) with a
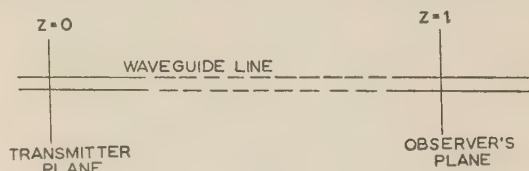
Fig. 1.—The co-ordinate system at time t.

'transmitter' at one point along its length (co-ordinate, $z = 0$). Suppose, further, that the field intensity in the waveguide is proportional to the signal intensity (linear system), then the function of the waveguide is to transform, in time (t) and distance (z), the signal—as produced by the transmitter $\Phi(t)$—into a 'distorted' signal $\phi(t)$. The ratio of the two functions, $\phi(t)/\Phi(t)$, will be referred to as the transfer function of the transmission system.

Clearly there is a reciprocal relation between the two functions of the type considered in the transform calculus, and the problem of determining the function $\phi(t)$ is solvable, at least in principle, using conventional methods of the transform calculus.

---

* For a discussion, see Stratton's 'Electromagnetic Theory', pp. 333–340.

The difficulty is not, however, that of principles but is of a practical nature and arises from two different causes. First, there is the difficulty of translating the physical properties of the waveguide into the language of mathematics, i.e. the determination of the transfer function T(ω, z). Secondly, there is an even greater difficulty of manipulating the mathematical expressions with a view to obtaining the function $\phi(t)$. In short, a truly exact solution to the problem is not possible except in a few isolated cases.

Nevertheless a judicious use of approximations does enable one to obtain asymptotic expansions for the $\phi(t)$ function, which can be made accurate to any desired degree, and this is the proposed method of approach.

(2.2) Some Elementary Problems

Let $\alpha(\omega)$ be the attenuation and $\beta(\omega)$ the phase characteristic of a waveguide. Evidently, if $\alpha(\omega)$ is a constant and $\beta(\omega)$ is a constant multiple of ω, the transfer function is given by

$$T(\omega, z) = \varepsilon^{-(\alpha+j\beta)z} \quad . \quad . \quad . \quad . \quad (1)$$

and the signal is transmitted without distortion, namely

$$\phi(t) = \varepsilon^{-(\alpha+j\beta)z}\Phi(t) \quad . \quad . \quad . \quad . \quad (2)$$

Thus, if a transmission system has a transfer function given by eqn. (1), all frequency components of the signal suffer equal attenuation and equal delay, and the signal transmission is distortion-free except for a decrease in amplitude.

In general, however, α is not a constant and β is not related linearly to frequency, and this fact *alone* is responsible for the signal distortion.

A transient phenomenon occupies, in principle, an infinite bandwidth, and consequently use is made of the whole of the α and β characteristics; for this reason, it is not possible to produce an exact solution. Sufficiently accurate results can, however, be obtained in many cases by assuming that most of the energy of the transient is transmitted over a relatively narrow band of frequencies, and then approximating to the α and β characteristics. To illustrate the method, consider the α and β characteristics of a 'perfect' waveguide (one whose walls are perfectly conducting).

Suppose that most of the transient energy is transmitted in a relatively narrow frequency band centred around ω_0 (Fig. 2), which is not too close to the cut-off frequency, ω_c. It is known that in this case α and β are given by

$$\alpha = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\beta = \sqrt{(k^2 - h^2)}$$

$$= \frac{1}{c}\sqrt{(\omega^2 - \omega_c^2)} \quad . \quad . \quad . \quad . \quad (4)$$

The *phase velocity* of the transient is given by the reciprocal of the slope of the line drawn from the origin to the operating point, A, on the β-characteristic, i.e.

$$(v_{ph})_0 = \frac{\omega_0}{\beta_0} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

It is essential to stress that this is so 'by definition' and that the upper and lower sidebands, on their own, have phase velocities which are somewhat lower and higher than $(v_{ph})_0$, respectively.

The phase velocity, as defined above, is somewhat fictitious, and though it has real physical meaning in narrow-band work it loses all significance if the signal occupies a very wide band of frequencies.
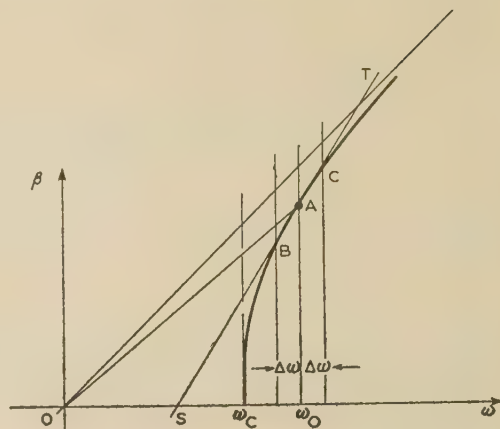
Fig. 2.—The β/ω characteristic of a perfect waveguide.

In any specific case, two sidebands of equal amplitude, placed one on each side of ω_0 and separated from it by $\frac{1}{2}\delta\omega$, combine to a signal given by

$$\exp j[(\omega_0 - \tfrac{1}{2}\delta\omega)t - (\beta - \tfrac{1}{2}\delta\beta)z]$$
$$+ \exp j[(\omega_0 + \tfrac{1}{2}\delta\omega)t - (\beta + \tfrac{1}{2}\delta\beta)z]$$
$$= 2\cos\tfrac{1}{2}(t\delta\omega - z\delta\beta)\exp j(\omega t - \beta z) \quad (6)$$

If the functional relation between ω and β is not linear, we can expand the β-characteristic around the point ω_0 in the Taylor's series as follows:

$$\delta\beta = \delta\omega\frac{d\beta}{d\omega} + \frac{(\delta\omega)^2}{2}\frac{d^2\beta}{d\omega^2} + \cdots \quad . \quad . \quad . \quad (7)$$

The first derivative, $d\omega/d\beta$, gives the velocity with which the intelligence of the signal is propagated (group velocity), while the second- and higher-order terms are the distortion terms. In particular, if the signal bandwidth is sufficiently small it is permissible to neglect all except the first two terms of the expansion (7). We then define ϕ by

$$2\phi = t\delta\omega - z\delta\beta$$
$$= \left(t - l\frac{d\beta}{d\omega}\right)\delta\omega - l\frac{(\delta\omega)^2}{2}\frac{d^2\beta}{d\omega^2} \quad . \quad . \quad . \quad (8)$$

and ϕ_0, the group delay, by

$$2\phi_0 = \left(t - l\frac{d\beta}{d\omega}\right)\delta\omega \quad . \quad . \quad . \quad . \quad (9)$$

The difference, $\phi_0 - \phi$, is responsible for the distortion of the intelligence and will be called the delay-distortion term. Clearly

$$2\Delta\phi = \phi - \phi_0 = \tfrac{1}{2}l(\delta\omega)^2\frac{d^2\beta}{d\omega^2}$$

or

$$\Delta\phi = l\left(\frac{\delta\omega}{2}\right)^2\frac{1}{\beta}\frac{1}{c^2}\left(\frac{h}{\beta}\right)^2 \quad . \quad . \quad . \quad (10)$$

Consequently the bandwidth, $\delta\omega$, of the intelligence transmitted is related to $\Delta\phi$ and the waveguide parameters by

$$\delta\omega = 2\left(\frac{\beta}{l}\right)^{1/2}\left(\frac{\beta}{h}\right)c(\Delta\phi)^{1/2} \quad . \quad . \quad (11)$$

The quantity $2\Delta\phi$ measures the time delay between the extreme sidebands after transmission through a waveguide length l, and as such is a measure of the waveguide distortion. For example

$2\phi = \pi/2$, an inversion of sidebands takes place: a condition of severe distortion.

For a given delay distortion, $2\Delta\phi$, eqn. (10) enables us to calculate the maximum bandwidth $\delta\omega$ which can be transmitted over a waveguide length l.

Undoubtedly, eqn. (10) gives the right functional relation between the maximum bandwidth a waveguide can handle and the length l, but unfortunately it is not easy to translate the magnitude of the quantity $2\Delta\phi$ into the nature and magnitude of transient distortion.

(3) THE TRANSMISSION OF A UNIT-STEP-MODULATED CARRIER

(3.1) Method of Analysis

It is assumed that the waveguide has a propagation coefficient, γ, given by

$$\gamma(p) = \alpha(p) + j\beta(p) \quad . \quad . \quad . \quad . \quad (12)$$

which is a function of the complex frequency, p. Further, it is assumed that the transmitted signal at $z = 0$ is $F(t)$ and has a frequency spectrum given by

$$f(p) = \int_0^\infty F(t)\varepsilon^{-pt}dt \quad . \quad . \quad . \quad . \quad (13)$$

If the system is linear, the output signal at a point distant z from the transmitter is given by the Carson's integral

$$S(z, t) = \frac{1}{2\pi j}\int_{C_p} f(p)\varepsilon^{pt - z\gamma(p)}dp \quad . \quad . \quad . \quad (14)$$

The contour of the integration, C_p, has the conventional form; it runs from $c - j\infty$ to $c + j\infty$, and the quantity c is chosen to suit the particular circumstances.

The transform $f(p)$ is usually easy to obtain, but there are a number of difficulties connected with the evaluation of the integral contained in eqn. (14).

For a perfect waveguide, $\gamma(p)$ is simply given by

$$\gamma(p) = \frac{1}{c}\sqrt{(p^2 + \omega_c^2)} \quad . \quad . \quad . \quad . \quad (15)$$

In the case of an imperfect waveguide, $\gamma(p)$ assumes a more complicated expression[4] which now depends on the shape of the waveguide and the type of mode supported within it, as well as on the surface impedance[5] of its walls, Z_s. The latter describes the physical property of the wall and is, in turn, a function of the frequency $[Z_s = Z_s(p)]$.

In the case of a unit-step-modulated carrier, the transform $f(p)$ is given by $1/(p - j\omega_0)$, so that, for a perfect waveguide, the signal is given by

$$S(z, t) = \frac{1}{2\pi j}\int_{C_p} \frac{\exp\left[pt - \frac{z}{c}\sqrt{(p^2 + \omega_c^2)}\right]}{p - j\omega_0}dp \quad . \quad . \quad (16)$$

The integrand has a pole at $p = j\omega_0$ and branch points at $p = \pm j\omega_c$, as shown in Fig. 3, and the path of integration runs along the imaginary axis with indentations as shown.

The integrand possesses a saddle point on the imaginary axis, as shown in Fig. 3, and under certain conditions it is possible to obtain asymptotic expansions of $S(z, t)$ by deforming the path of integration through the saddle point along the path of steepest descents. The method cannot, however, be readily generalized and requires, at times, a considerable amount of justification.

On the other hand, a method of obtaining asymptotic expansions, as developed by Van der Waerden,[6] is readily adapted for our immediate purpose.
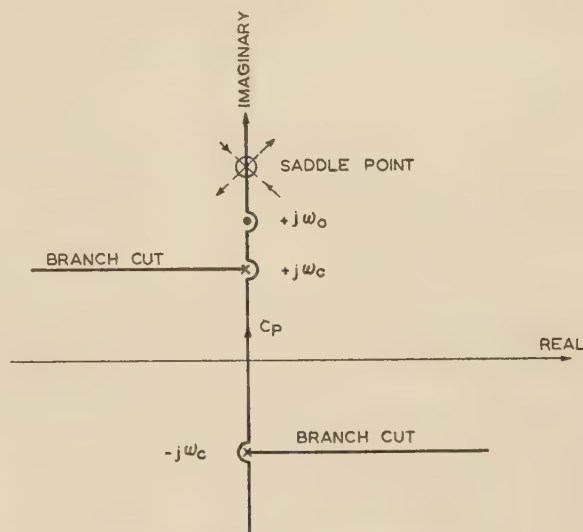
Fig. 3.—The C_p contour in the p-plane.

To that end we make the substitution

$$u(p) = -p + V\sqrt{(p^2 + \omega_c^2)} \quad . \quad . \quad (17)$$

where

$$V = \frac{z/t}{c} \quad . \quad . \quad . \quad . \quad . \quad (18)$$

and transform the integral to the u-plane, giving

$$S(z, t) = \frac{1}{2\pi j}I_u \quad . \quad . \quad . \quad . \quad (19)$$

where

$$I_u = \int_{C_u} P(u)\varepsilon^{-tu}du \quad . \quad . \quad . \quad (20)$$

with

$$P(u) = \frac{1}{p - j\omega_0}\frac{dp}{du} \quad . \quad . \quad . \quad (21)$$

The function $P(u)$ is characterized by a branch point at

$$u_B = -j\omega_c(1 - V^2)^{1/2} \quad . \quad . \quad . \quad (22)$$

and a pole at

$$u_p = j[-\omega_0 + V\sqrt{(\omega_0^2 - \omega_c^2)}] \quad . \quad . \quad (23)$$

Consequently, unless the pole is in the immediate vicinity of the branch cut, the integral can be conveniently split into two parts, I_P and I_B, as contributed by the two contours, C_P and C_B, respectively (as shown in Fig. 4).
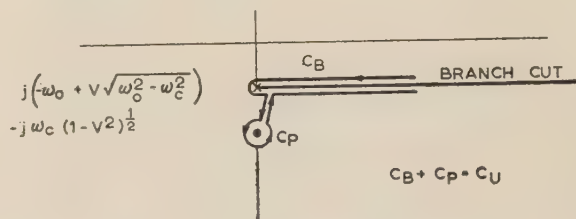
Fig. 4.—The C_u-contour in the u-plane.

It will be found later that the signal $S(z, t)$ can conveniently be considered to be composed of three states in time: (a) a region of *antecedents* (these travel with a velocity exceeding the group velocity), which precedes (b) a steep wavefront, travelling at the group velocity, followed by (c) the region of *posterior transients*. The antecedent transients comprise the *B-precursor* and the *P-precursor*, and in essence are the asymptotic expansions of the integrals I_B and I_P, for $V \to 1$.

(3.2) The Asymptotic Solution for the Antecedents

(3.2.1) The Perfect Waveguide.

As explained above we are seeking an asymptotic expression for I_B given by

$$I_B = \int_{C_B} P(u)\varepsilon^{-tu}du \quad . \quad . \quad . \quad . \quad (24)$$

where C_B runs around the branch cut, as shown in Fig. 4.

Through the substitution

$$s = (u - u_B)^{1/2} \quad . \quad . \quad . \quad . \quad (25)$$

we pass into the s-plane and expand $P(u)$ in ascending powers of s,

$$P(u) = b_{-1}s^{-1} + b_0 + b_1 s + \dots \quad . \quad . \quad (26)$$

The series leads to an asymptotic solution to eqn. (24), and for large enough values of t it is sufficiently accurate to retain the leading term only, provided, of course, that the pole is not in the immediate vicinity of the branch cut.

Using eqn. (21) it can be shown that

$$b_{-1} = \frac{\varepsilon^{j\gamma/4}}{\sqrt{2}} \frac{V\sqrt{\omega_c}}{[\omega_0\sqrt{(1-V^2)} - \omega_c](1-V^2)^{1/4}} \quad . \quad (27)$$

For $V \to 0$,

$$b_{-1} \to \frac{\varepsilon^{j\pi/4}}{\sqrt{2}} \frac{V\sqrt{\omega_c}}{\omega_0 - \omega_c} \quad . \quad . \quad . \quad (28)$$

while for $V \to 1$,

$$b_{-1} \to -\frac{\varepsilon^{j\pi/4}}{\sqrt{2}} \frac{1}{\sqrt{\omega_c}(1-V^2)^{1/4}} \quad . \quad . \quad (29)$$

Further, the substitution, eqn. (26), converts eqn. (24) into a series of fractional powers of t. Thus, retaining the first term only, we have

$$I_B \sim 2\sqrt{(\pi)}b_{-1}t^{-1/2}\exp[j\omega_c t\sqrt{(1-V^2)}] \quad . \quad . \quad (30)$$

Consequently the transient is

$$S(z,t)_B \sim \frac{\varepsilon^{-j\pi/4}}{\sqrt{(2\pi)}} \frac{V\sqrt{\omega_c}}{\omega_0 - \omega_c}t^{-1/2}\exp[j\omega_c t\sqrt{(1-V^2)}] \quad (31)$$
$$V \to 0$$

and

$$S(z,t)_B \sim \frac{\varepsilon^{-j3\pi/4}}{\sqrt{(2\pi)}} \frac{1}{\sqrt{\omega_c}\left[t^2 - \left(\frac{z}{c}\right)^2\right]^{1/4}}$$
$$\exp\left\{j\omega_c\sqrt{\left[t^2 - \left(\frac{z}{c}\right)^2\right]}\right\} \quad . \quad (32)$$
$$V \to 1$$

For sufficiently small values of V, and for V sufficiently close to unity, expressions (31) and (32) become, respectively,

$$S(z,t)_B \sim \frac{\varepsilon^{-j\pi/4}}{\sqrt{(2\pi)}} \frac{\sqrt{\omega_c}}{\omega_0 - \omega_c}\frac{z}{c}t^{-3/2}\varepsilon^{j\omega_c t} \quad . \quad (33a)$$
$$= P_B$$

$$S(z,t)_B \sim \frac{\varepsilon^{j3\pi/4}}{\sqrt{(2\pi)}} \frac{\varepsilon^{j\tau^{1/4}}}{\tau^{1/4}} \quad . \quad . \quad . \quad . \quad (33b)$$
$$= A_B$$

where

$$\tau = \omega_c^2\frac{z}{c}\left(t - \frac{z}{c}\right) \quad . \quad . \quad . \quad . \quad (34)$$

Eqn. (31) [or (33a)] represents one of the posterior transients, while eqn. (33) [or (32b)] represents one of the antecedent transients, which we shall call the 'B-precursor' (Fig. 5).

We note the distinct nature of the P_B (B-sequel) and A_B
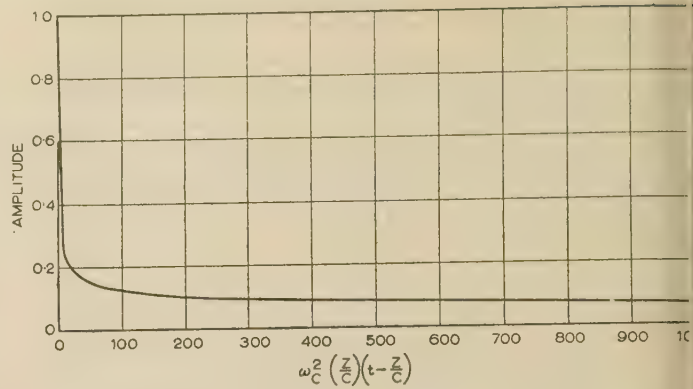
Fig. 5.—The B-precursor as a function of time, t.

(B-precursor) transients. P_B has an effective frequency* which is a decreasing function of time and which tends to ω_c; the amplitude of the P_B transient is proportional to the $-3/2$ power of time. The virtual frequency of the B-precursor, however, is infinite for $t = z/c$ and is a decreasing function of time; its amplitude is proportional to the $-1/4$ power of time.

(3.2.2) The Imperfect Waveguide.

The analysis for an imperfect waveguide differs from that of the perfect one in that $\gamma(p)$ is related to frequency, not in the simple manner indicated in eqn. (15) but by a more complicated expression[4] involving, among other factors, the surface impedance Z_m as a function of the complex frequency, p.

Since the B-precursor is transmitted by a carrier much in excess of ω_0, it is therefore permissible to make certain approximations, simplifying the analysis, and it can be shown, for example, that in the case of a plain metallic perfectly smooth circular waveguide supporting any H_0-mode, the B-precursor has essentially the same form as given by eqn. (32a), except for an attenuating factor.

Similar expressions can be derived for other modes and other waveguide shapes. There is little significance, however, in quoting the formulae—although the shape of the B-precursor is largely affected by the waveguide shape and the mode carried within it—because the information is of academic interest only and has no practical significance whatever, for the following reason.

As mentioned above, the B-precursor is transmitted on a carrier whose frequency is extremely high (much in excess of ω_0), and at very high values of frequency it is no longer true that a metal surface has a normalized surface impedance simply given by (see remarks in Reference 4)

$$Z_m = \sqrt{\left(\frac{p\mu_m}{\sigma}\right)}\sqrt{\frac{\epsilon}{\mu_0}} \quad . \quad . \quad . \quad . \quad (35)$$

On the contrary, the surface impedance of a metal surface tends, at these very high frequencies, to be dominated by the physical state of the surface rather than by the electrical properties of the metal itself—and thus, as mentioned elsewhere,[4] it is totally unpredictable.

(3.3) The Principal Part of the Signal

(3.3.1) The Perfect Waveguide.

In the case of the perfect waveguide, eqns. (17) to (23) are relevant. We seek a solution to eqn. (20) as a function of the parameter V, and we note that as V varies from 1 to 0 the pole and the branch point move relative to each other in the complex plane of u, and care must be exercised when expanding $P(u)$ in

* Instantaneous frequency is implied.

infinite series, to avoid introducing any singularities originally
t present in the integrand and to be able to follow the passage
the pole unambiguously.

We change the variable from u to v through the substitution

$$v = u - u_B \qquad \ldots \quad \ldots \qquad (36)$$

that in the v-plane we have a branch point at

$$v_B = 0 . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

d a pole at

$$v_P = j[-\omega_0 + V(\omega_0^2 - \omega_c^2)^{1/2} + \omega_c(1 - V^2)^{1/2}] . \quad (38)$$

onsequently we find that

$$\left.\begin{array}{l} (v_P)_{V=0} = j(-\omega_0 + \omega_c) \\ (v_P)_{V=1} = j[-\omega_0 + (\omega_0^2 - \omega_c^2)^{1/2}] \end{array}\right\} \quad . \quad . \quad (39)$$

hile for $V = \sqrt{(1 - (\omega_c/\omega_0)^2)}$ the pole and the branch point
incide.

It is, however, essential to observe that for $V = 1$ the pole is
the 'wrong'* Riemannian plane. The passage of the pole is
arked in Fig. 6, but for the sake of clearness the poles for
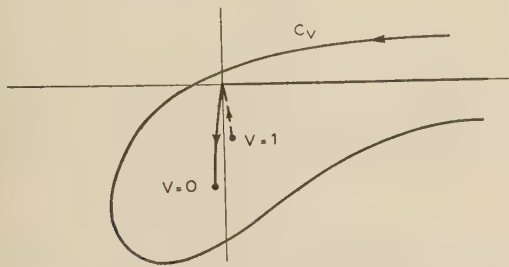
Fig. 6.—The passage of the pole in the complex plane of v.

$= 0$ and $V = 1$ have been placed slightly to each side of the
maginary axis, while in actual fact the pole is always on the
maginary axis.

The original integral, I_u [eqn. (20)], has now been transformed
to one in the v plane, namely

$$I_u = \varepsilon^{-tu_B} I_v$$

here

$$I_v = \int_{C_v} P(u)\varepsilon^{-tv} dv \quad . \quad . \quad . \quad . \quad (40)$$

The path of integration, C_v, is as shown in Fig. 6.

Further, the substitution

$$s = \sqrt{v} \quad . \quad . \quad . \quad . \quad . \quad : \quad (41)$$

ransforms I_v into an integral (in the plane of s) whose integrand
free from branch points but has a pole as indicated in Fig. 7.
he path of integration, C_s, can be distorted, as shown in Fig. 7.
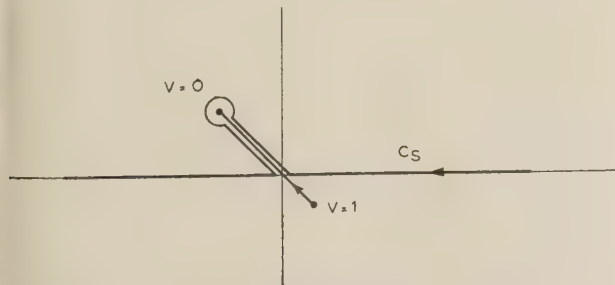
ig 7.—The position of the pole in relation to the integration contour
C_s in the plane of s.

* By 'wrong' we mean the upper leaf of the Riemannian plane if the integration
ath is confined to the lower plane, and vice versa.

The integral I_v now becomes

$$I_v = 2\int_{C_s} sP(u)\varepsilon^{-ts^2} ds \quad . \quad . \quad . \quad (42)$$

The integrand is expandable in its Laurent series around the
pole, and the pole in question is given by

$$b = \sqrt{v_P} \quad . \quad . \quad . \quad . \quad . \quad (43)$$

Thus $\qquad sP(u) = \dfrac{a_{-1}}{s - b} + a_0 + a_1(s - b) + \ldots \quad . \quad (44)$

With this substitution, eqn. (42) yields an asymptotic expansion
for I_v, and its leading term gives the required solution to the
problem, which is

$$S(z, t)_P \sim \varepsilon^{-tu_P} \frac{1}{\sqrt{\pi}} \operatorname{erfc}(jb\sqrt{t}) \quad . \quad . \quad (45)$$

where $\qquad \operatorname{erfc}(x) = \int_x^\infty \varepsilon^{-y^2} dy \quad . \quad . \quad . \quad . \quad (46)$

For large values of the argument, we can use the asymptotic
expressions for erfc (x), so that, if $b\sqrt{t}$ is a large quantity, we have

$$S(z, t)_P \sim \left(1 - \frac{1}{\sqrt{\pi}} \frac{\varepsilon^{b^2 t}}{2jb\sqrt{t}}\right)\varepsilon^{-tu_P} \quad . \quad . \quad (47)$$

if b lies above the real axis

and $\qquad S(z, t)_P \sim \frac{1}{\sqrt{\pi}} \frac{\varepsilon^{-b^2 t}}{2b\sqrt{t}} \varepsilon^{-tu_P} \quad . \quad . \quad . \quad (48)$

if b lies below the real axis.

Eqn. (48) represents the P-precursor while eqn. (47) represents
a posterior transient (P-sequel). The reason for this nomen-
clature will be explained later.

If the amplitude of the transmitted unit-step-modulated carrier
is unity, then, according to eqns. (45) to (48), at a point distant
z from the transmitter there will be no signal received until a
time $t_0 = z/c$ when the P-precursor [given by eqn. (48)] arrives.
This quickly grows in amplitude until at a time $t_g = z/v_g$
($v_g =$ group velocity) it reaches one-half the original amplitude.
Thereafter the amplitude of the signal oscillates about the mean
value equal to unity, to which value it settles after a long time.

The time t_0 is of importance in that it marks the time of the
arrival of the P-precursor. The time t_g is of importance because
it is the *virtual time* of signal travel.

It is of interest to note that the slope of the received signal
at $t = t_g$ is given by

$$\frac{\sqrt{\pi}}{T} = \left(\frac{\omega_0}{2} \frac{c}{z}\right)^{1/2} \frac{\omega_0}{\omega_c}\left[1 - \left(\frac{\omega_c}{\omega_0}\right)^2\right] = A \quad . \quad (49)$$

and this, apart from a constant factor, is the same as eqn. (11),
provided that the waveguide is operated so that ω_c/ω_0 is a small
number. One would intuitively expect this agreement to take
place.

The rise time T at $t = t_g$ is related to the bandwidth-handling
capacity of the waveguide, but the relation is by no means as
simple as in the case of an ideal i.f. amplifier whose bandwidth
is equal to one-half the reciprocal of the rise time.

The response of a length of waveguide to unit-step-modulated
carrier is shown in Fig. 8 as a function of At_1, where $t_1 = t - t_g$.
The curve has been plotted [using eqn. (45) and existing
Tables[7,8,9]] against At_1 rather than t_1 because the curve so
obtained is a universal one, applicable to all waveguides under
all conditions.
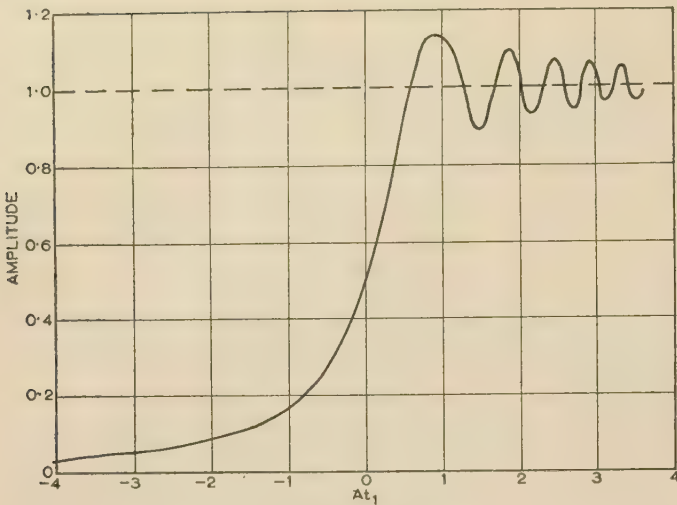
Fig. 8.—The transient response of a waveguide to the unit-step-modulated carrier.

(3.3.2) **The Imperfect Waveguide.**

In the case of an imperfect waveguide, the propagation coefficient $\gamma(p)$ is given by[4]

$$\gamma = [h^2 - k^2 + 2MjZ_s(p)]^{1/2}$$

where the quantity M is a function of frequency, waveguide geometry and mode.

Consequently, for sufficiently small values of surface impedance, we may approximate and take instead of the u function given by eqn. (17) the following:

$$u = -p + V\sqrt{[p^2 + \omega_c^2 + 2(jR - X)]} \qquad . \qquad (50)$$

where R and X are respectively proportional to the real and imaginary parts of the surface impedance and are essentially small quantities.

The path of integration in the p-plane runs as shown in Fig. 9.
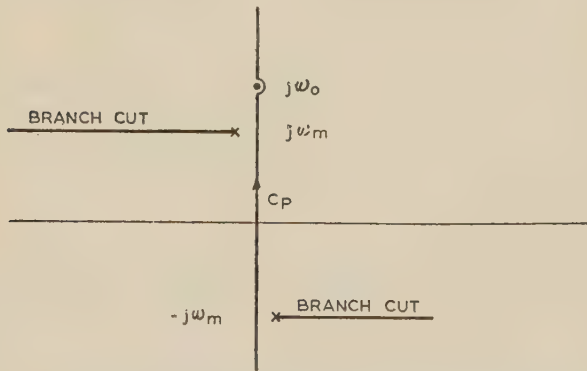
Fig. 9.—The p-plane pertaining to an imperfect waveguide.

This Figure also shows the distribution of poles and branch points.

If u_B is the branch point, the substitution $v = u - u_B$ transforms the integral into the v-plane, and Fig. 10 shows the passage of the pole as V varies from $V = 1$ to $V = 0$.

The contour integral is solved in the manner indicated in Section 3.3.1, and the following solution can be obtained:

$$S(z, t)_P \sim \left[1 - \frac{1}{\sqrt{\pi}} \operatorname{erfc}(jb\sqrt{t}) \right] \varepsilon^{-tu_P} \qquad . \qquad . \quad (51)$$
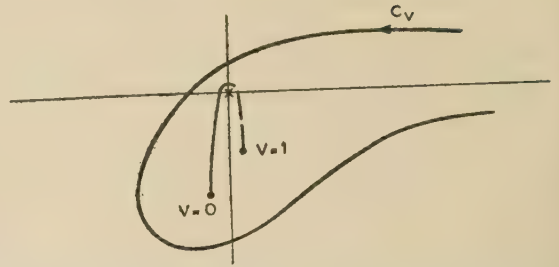
$$\text{for } \mathcal{R}(b^2) \leqslant 0$$

Fig. 10.—The passage of the pole in the plane of v for an imperfect waveguide.

$$S(z, t)_P \sim \frac{1}{\sqrt{\pi}} \operatorname{erfc}(b\sqrt{t})\varepsilon^{-tu_P} \qquad . \qquad . \qquad (52)$$

$$\text{for } \mathcal{R}(b^2) > 0$$

where

$$b^2 = j\left[-\omega_0 + V\sqrt{(\omega_0^2 - \omega_c^2)} - V\frac{jR - X}{\sqrt{(\omega_0^2 - \omega_c^2)}} \right]$$

$$+ j\left(\omega_c + \frac{jR - X}{\omega_c} \right)(1 - V^2)^{1/2} \quad (53)$$

and $\quad u_P = j\left\{ -\omega_0 + V\sqrt{(\omega_0^2 - \omega_c^2)}\left[1 - \frac{jR - X}{\sqrt{(\omega_0 - \omega_c^2)}} \right] \right\} \quad (54)$

Eqns. (51) and (52) show that the response is essentially the same as that of a perfect waveguide, except for an attenuating factor.

(3.4) **Discussion of the Results**

With the help of times t_0 $(=z/c)$ and t_g $(=z/v_g)$, as introduced in Section 3.3, we can sketch the response of a waveguide of length l $(z = l)$, to a unit-step-modulated carrier, as shown in Fig. 11, which, of course, is not drawn to scale.

Until time t_0 no signal is received. Time t_0 marks the beginning of a region of weak field intensity, which we propose to call the region of antecedent transients. Antecedents are made up of two precursors: the B-precursor and the P-precursor. The B-precursor is the asymptotic expansion of the branch cut integral I_B, for $V \to 1$ [eqn. (33)], while the P-precursor is the asymptotic expansion of the pole integral I_P, also for $V \to 1$ [eqn. (48)]. The two precursors are distinct in nature, and their relative magnitudes are plotted in Fig. 12. Except in the immediate vicinity of t_g, the B-precursor is (for a perfect waveguide) much larger than the P-precursor.

In the immediate vicinity of the time t_g the rate of rise of the field strength with time is very large, and the field soon reaches intensities of the order of unity, the rate of rise being inversely proportional to the quantity A (eqn. 49), i.e. among other factors to the square root of the distance, l.

The time t_g marks the beginning of a region in which the mean field strength is approximately unity with transients superimposed upon it. Here again there are two posterior transients: the B-sequel and the P-sequel. The B-sequel is represented by the asymptotic expansion of the branch cut integral I_B for $V \to 0$ [eqn. (31)], while the P-sequel is given by the oscillating part of the asymptotic expansion of the pole integral I_P for $V \to 0$ [the second term in the brackets of eqn. (47)].

The signal in the immediate vicinity of the time t_g is plotted in Fig. 8.

Fig. 11 shows, in addition, the variation of the following quantities: virtual carrier frequency, ω_v, virtual phase velocity, $(v_{ph})_v$, and virtual group velocity, $(v_g)_v$. These quantities are instantaneous values and have the following meaning.

The envelope of the signal is effectively transmitted on a
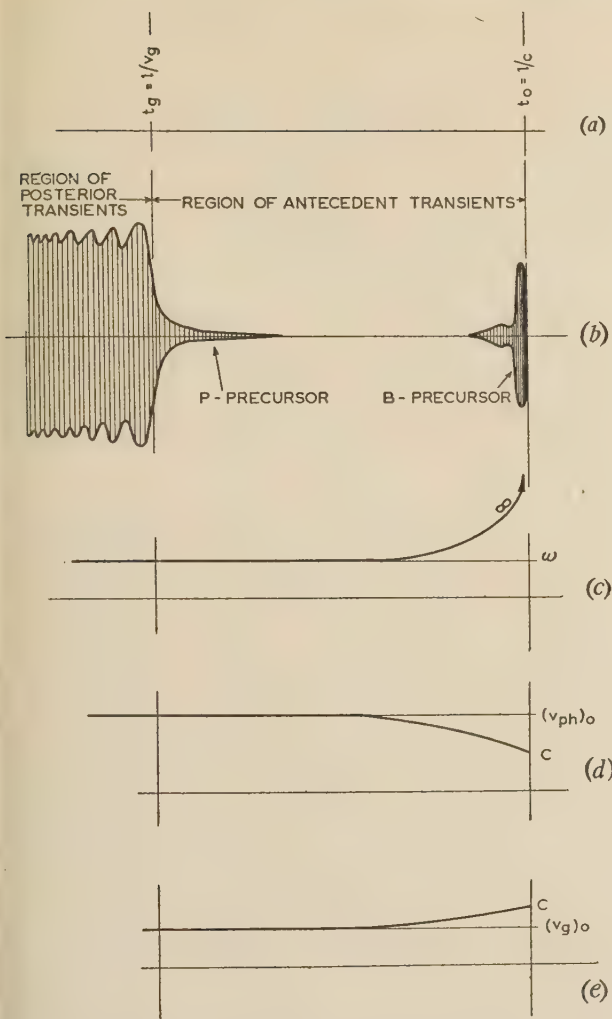
Fig. 11.—The response of a waveguide to a unit-step-modulated carrier.

(a) Time scale at a point distant l from the transmitter.
(b) Response function.
(c) Virtual carrier frequency, ω_v.
(d) Virtual phase velocity, $(v_{ph})_v$.
(e) Virtual group velocity, $(v_g)_v$.
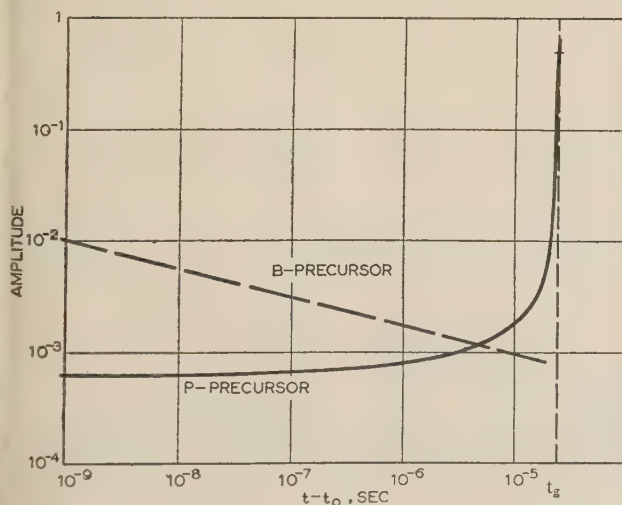
Fig. 12.—The relative magnitudes of the B- and P-precursors.

Waveguide constants:
$l = 30$ km.
$f_0 = 10$ kMc/s.
$f_c/f_0 = 0\cdot6$.

carrier whose frequency changes as the time increases. At the time t_0, the instantaneous frequency of the carrier is infinite, but it falls rapidly, as the time increases, to the value ω_0, as shown in Fig. 11.

Similarly, the effective phase velocity of the carrier is at time t_0 equal to c, but as the time increases this velocity also increases and finally settles down to the value $(v_{ph})_0$.

Conversely, any point on the envelope appears to move with a velocity that decreases with time from the value c at t_0 to the asymptotic value v_g. For this reason, the envelope of a signal passing down the line appears to 'stretch' in time, as the time increases.

(4) THE PROPAGATION OF R.F. PULSES IN WAVEGUIDES

It is convenient to make two simplifying assumptions, although these have little bearing on the shape of pulses likely to be met in practical applications.

In the first place, the B-precursor and the B-sequel can be neglected. Secondly, it can be assumed that the pulse is transmitted on a carrier, ω_0, sufficiently high in frequency in relation to the pulse width to ensure that in the pulse interval there are a large number of cycles of microwave energy (say 10 or more), so that the random phasing effects of the carrier relative to the pulse edges may be neglected.

With these assumptions, the envelope of the unit-step-modulated carrier at a distance z and time t is given by [eqn. (45)]

$$\Lambda(t, z) = \frac{1}{\sqrt{\pi}} \operatorname{erfc}(jb\sqrt{t}) \quad . \quad . \quad . \quad (55)$$

where b is given by

$$b^2 = j[-\omega_0 + V(\omega_0^2 - \omega_c^2)^{1/2} + \omega_c(1 - V^2)^{1/2}] \quad (56)$$

with

$$V = \frac{z/t}{c}$$

Since the principle of superposition is applicable, the response to a pulse of duration Δt is given by

$$\Psi_1(t, z, \Delta t) = \Lambda(t, z) - \Lambda(t + \Delta t, z) \quad . \quad . \quad (57)$$

Again, it will be convenient to introduce the quantity A, given by eqn. (49) as

$$A = \left(\frac{\omega_0}{2}\frac{c}{z}\right)^{1/2}\frac{\omega_0}{\omega_c}\left[1 - \left(\frac{\omega_c}{\omega_0}\right)^2\right]$$

and then Λ [eqn. (55)] becomes the normalized complex Fresnel integral of A and time t_1 defined by

$$t_1 = t - t_g \quad . \quad . \quad . \quad . \quad . \quad (58)$$
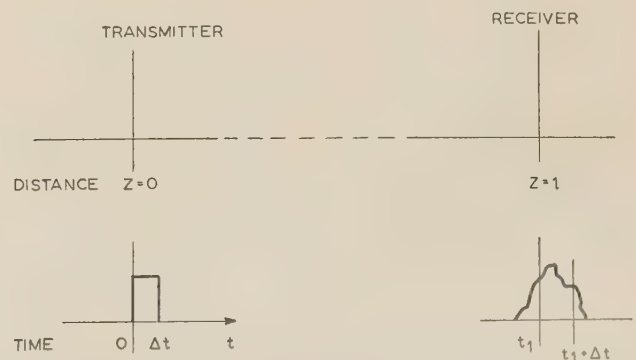
Fig. 13.—Time and space co-ordinates of a transmitted and received transient.

Time $t_1 = l/v_g$.
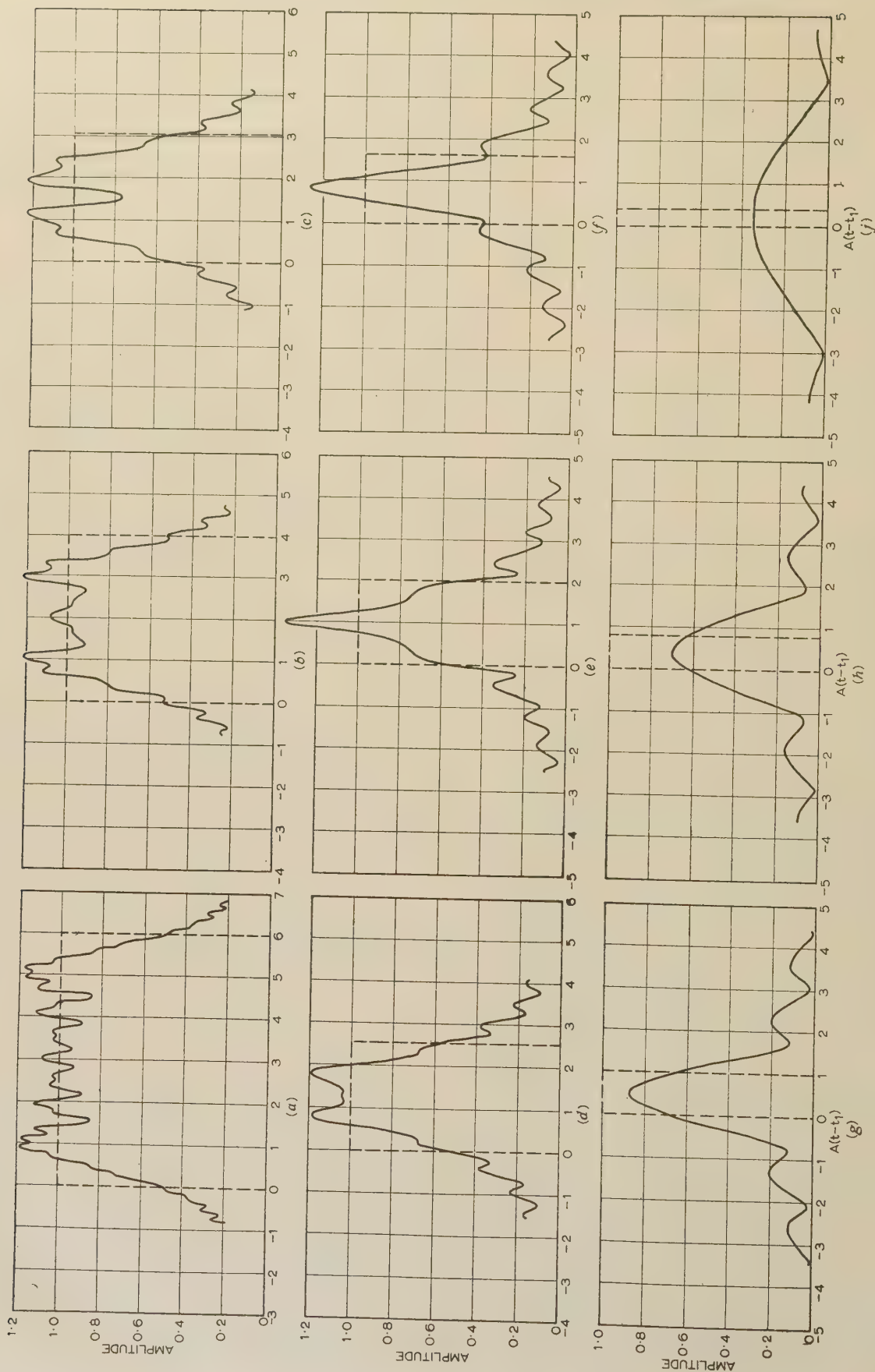
Fig. 14.—Envelope of the received pulse after transmission through a waveguide.

(a) $A\Delta t = 6.0$.　　(b) $A\Delta t = 4.0$.　　(c) $A\Delta t = 3.04$.
(d) $A\Delta t = 2.56$.　　(e) $A\Delta t = 2.0$.　　(f) $A\Delta t = 1.6$.
(g) $A\Delta t = 1.0$.　　(h) $A\Delta t = 0.8$.　　(j) $A\Delta t = 0.4$.

If now we introduce the time and space co-ordinates as indicated in Fig. 13 we find that the response function is given by

$$\Psi_1(t, z, \Delta t) = \Psi(At_1, A\Delta t) \quad . \quad . \quad . \quad (59)$$

The function Ψ is shown plotted in Fig. 14 against At_1 for the different values of the parameter $A\Delta t$. From this Figure the general deterioration of the pulse shape as the parameter $A\Delta t$ is decreased is evident.

For large values of $A\Delta t$ the pulse retains much of its original shape, but its rising and falling edges have finite slopes determined by the quantity At_1. Gradually, however, as $A\Delta t$ decreases, the pulse begins to disperse and suffer greater and greater distortion until, for values of $A\Delta t$ less than about unity, the pulse becomes dispersed beyond recognition.

It is stressed that the pulse distortion as introduced by the waveguide is basically of a nature different from that introduced by a perfect band-pass filter (e.g. an i.f. amplifier). Whereas, in the latter case, the information contained in the high-frequency end of the envelope spectrum is cut off by the filter and thereby lost irrecoverably, in the case of a waveguide the distortion of the pulse is in the nature of delay distortion and as such is removable (at least in principle) through the use of suitable equalizers.

(5) THE PROPAGATION OF A DOUBLE PULSE

To an engineer, Fig. 8 is of far greater assistance in estimating the transient behaviour of a waveguide than the simple formula [eqn. (11)], although the two have a little in common.

From Fig. 8, however, it is not immediately obvious what is the response of a waveguide to a c.w. pulse of microwave energy, and that is where the results from Section 4, together with Fig. 14, are of direct assistance.

From the point of view of bandwidth-handling capacity, however, we still have not obtained quite enough information. The question to which a transmission engineer seeks an answer is: what is the maximum rate at which various 'bits' of information can be transmitted and resolved at the receiving end without ambiguity?

We can answer the above question by transmitting two pulses separated by a small interval of time and discovering the maximum distance at which they can be received unambiguously as two separate pulses before they merge into one another. To that end it can be assumed that the transmitted signal is composed of two pulses each of duration Δt and separated from each other in time by Δt. The response of a length of waveguide to such a signal is evidently given by

$$\Omega[At_1, A\Delta t, A\Delta t] = \Psi(At_1, A\Delta t) - \Psi[A(t_1 + \Delta t), A\Delta t]$$
$$. \quad . \quad . \quad . \quad (61)$$

It can be shown that there is a fairly well-defined transition point after which the two pulses become indistinguishable. This occurs around $A\Delta t \simeq 1$.

The function Ω is shown plotted in Fig. 15, and it will be observed that, whereas for $A\Delta t = 1 \cdot 6$ the two pulses are distinct, for $A\Delta t = 0 \cdot 8$ it is no longer possible to interpret the signal as two distinct pulses.

(6) NUMERICAL EXAMPLES

It will be convenient to compare the bandwidth-handling capacity of waveguides per 30 km length ($\simeq 10^6$ ft). The figures for any other lengths can be easily obtained if it is remembered that the bandwidth capacity of a waveguide is inversely proportional to the square root of its length.

Example 1.

Consider a waveguide operated at $f_0 = 10$ kMc/s and having $\omega_c/\omega_0 = 0 \cdot 6$.
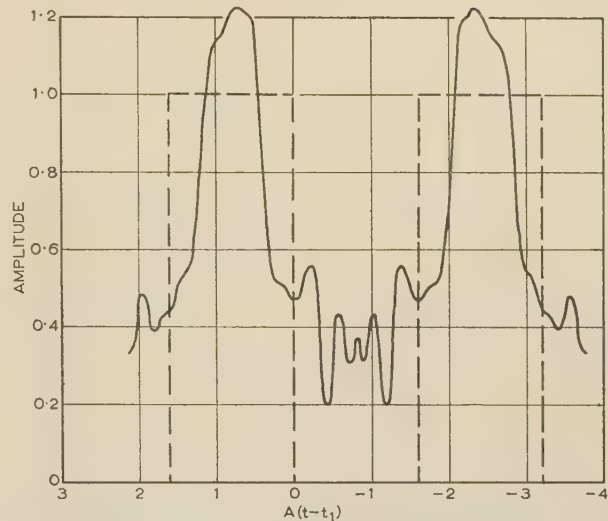
Fig. 15.—Received double pulse for $A\Delta t = 1 \cdot 6$.
The two pulses spaced by the pulse width.

Substituting this information in eqn. (11), we have

$$\frac{\Delta f}{f_0} = \sqrt{2}(\Delta\phi)^{1/2}\left(\frac{10^4}{2\pi 10^{10}}\right)^{1/2}\frac{1}{0 \cdot 6} \quad . \quad . \quad (62)$$

i.e.
$$\Delta f = 9 \cdot 37(\Delta\phi)^{1/2} \text{ Mc/s} \quad . \quad . \quad . \quad (63)$$

In other words, the waveguide can handle a $9 \cdot 37$ Mc/s bandwidth per $(1 \text{ rad})^{1/2}$ delay between the extreme ends of the band.

This result is not easy to interpret but it gives some idea of the orders of magnitude involved.

The quantity A, given by eqn. (49), is

$$A = (\pi 10^{10} 10^4)^{1/2}\left(\frac{1}{0 \cdot 6} - 0 \cdot 6\right) \quad . \quad . \quad (64)$$

Consequently, the slope of the response curve at time t_1 (see Fig. 8) is

$$S_0 = \frac{A}{\sqrt{\pi}} = 1 \cdot 07 \times 10^7 \quad . \quad . \quad . \quad (65)$$

so that the rate of rise of the field at this point is

$$T_0 = \frac{1}{S_0} = 93 \cdot 5 \text{ millimicrosec} \quad . \quad . \quad (66)$$

The rate of rise of the field is somewhat more rapid at a slightly later time (see Fig. 8), and is

$$T_{min} = 58 \text{ millimicrosec} \quad . \quad . \quad . \quad (67)$$

As far as the pulse behaviour of the waveguide is concerned, we turn to Fig. 14. Although the results may appear at first to be somewhat bewildering and disconcerting, the choice of a suitable value of the quantity $A\Delta t$ is quite simple, given the necessary information about the equipment working in conjunction with the waveguide. For example, neglecting the signal/noise requirements and band-limiting equipment outside the waveguide, it is apparent that the quantity $A\Delta t$ should not be smaller than about unity, but the choice is by no means critical. We may choose $A\Delta t = \sqrt{2}$, for which value we get

$$\Delta t = \frac{\sqrt{2}}{1 \cdot 9} \times 10^{-7} = 74 \text{ millimicrosec} \quad . \quad (68)$$

i.e. the waveguide is capable of transmitting recognizable pulses whose width is not smaller than 74 millimicrosec.
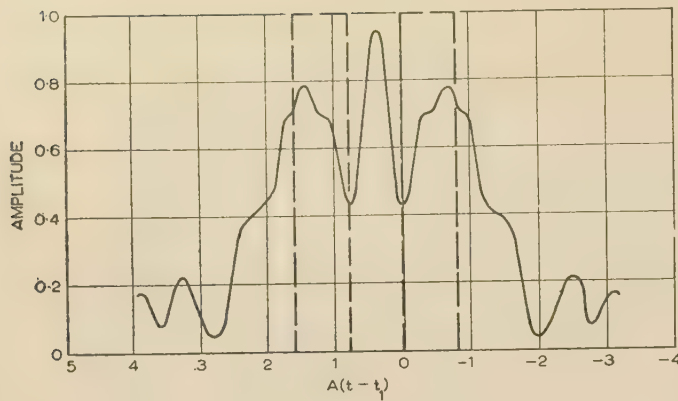
Fig. 16.—Received double pulse for $A\Delta t = 0 \cdot 8$.

The two pulses spaced by the pulse width.

At the time earlier than z/c no signal is received, but imme-
diately after this time the two precursors arrive; however, the
B-precursor is much larger in amplitude than the P-precursor,
though gradually the P-precursor grows while the B-precursor
decreases in amplitude.

After the arrival of the signal at the time z/v_g, the signal
intensity soon settles to a constant value with posterior transients
superimposed on it. These transients can conveniently be split
into the B-sequel and the P-sequel, which are physically distinct
and are mathematically the asymptotic expansions of the I_B
integral and the variable term of the asymptotic expansion of the
I_P integral, respectively.

It has been shown that the virtual carrier frequency, phase
velocity and group velocity are subject to large variations in the
time interval z/c to z/v_g.

The analysis of transients has been concluded by an application
of the theory to pulse propagation in waveguides, for which case
detailed graphical solutions have been obtained.

Further, study of Figs. 15 and 16 will reveal that under similar
conditions two pulses 74 millimicrosec wide and separated by
74 millimicrosec would just about be conveniently resolvable.

Example 2.

Consider a low-loss overmoded circular waveguide operated at
35 kMc/s with $\omega_c/\omega_0 = 0 \cdot 15$.

Using a notation as above, we find that for this waveguide

$$\left.\begin{array}{l} \Delta f = 70 \cdot 2 \times (\Delta\phi)^{1/2} \text{ Mc/s} \\ A = 21 \cdot 5 \times 10^7 \\ S_0 = 0 \cdot 121 \times 10^9 \\ T_0 = 8 \cdot 26 \text{ millimicrosec} \\ T_{min} = 5 \cdot 17 \text{ millimicrosec} \\ \Delta t = \dfrac{\sqrt{2}}{A} = 6 \cdot 7 \text{ millimicrosec} \end{array}\right\} \quad . \quad . \quad . \quad (69)$$

Thus, ideally, the waveguide could transmit resolvable pulses as
narrow as $6 \cdot 7$ millimicrosec and at intervals of $6 \cdot 7$ millimicrosec.

Among others, these examples illustrate the good agreement
(to the order of magnitude) between the results obtained from the
rigorous theory and the delay distortion formulae [eqn. (11)].
Thus, in the first example, if we say that the waveguide can carry
pulses not narrower than 74 millimicrosec, this can be interpreted,
in the light of the transmission through conventional ideal band-
pass elements, that the cut-off bandwidth is $1/(2\Delta f) = 6 \cdot 75$ Mc/s,
which is in reasonable agreement with $f = 9 \cdot 37 (\Delta\phi)^{1/2}$ Mc/s, as
given by eqn. (61). But the bandwidth limitation is of the delay-
distortion nature and is not due to the higher harmonics being
cut-off by the transmission medium.

(7) CONCLUSIONS

A theory of the propagation of transients in waveguides has
been developed.

The theory has been subsequently applied to a detailed investi-
gation of the propagation of a unit-step-modulated carrier and
the results for this case are summarized in a qualitative manner
in Fig. 11. It has been shown that the received signal is mathe-
matically composed of two terms: the pole integral I_P and the
branch cut integral I_B. Physically, the pole integral is the
principal part of the signal at the time z/v_g and thereafter. This
part of the signal is preceded by a region of antecedent transients
which can be resolved into the P-precursor and the B-precursor,
which are the asymptotic solutions of the I_P and I_B integrals,
respectively.

(8) ACKNOWLEDGMENTS

(9) REFERENCES

(1) MILLER, S. E., and BECK, A. C.: 'Low-loss Waveguide
      Transmission', *Proceedings of the Institute of Radio
      Engineers*, 1953, **41**, p. 348.
(2) BECK, A. C.: 'Waveguide Investigation with Millimicro-
      second Pulses', *Bell System Technical Journal*, 1956, **35**,
      p. 35.
(3) MILLER, S. E.: 'Waveguide as a Communication Medium',
      ibid., 1954, **33**, p. 1209.
(4) KARBOWIAK, A. E.: 'Exact Formulae and Curves for the
      Attenuation and Phase Propagation Coefficients in Wave-
      guides', Technical Memo. No. 140, Standard Telecom-
      munication Laboratories Limited.
(5) KARBOWIAK, A. E.: 'Theory of Imperfect Waveguides:
      Effect of Wall Impedance', *Proceedings I.E.E.*, Paper
      No. 1841 R, September, 1955 (**102 B**, p. 698).
(6) VAN DER WAERDEN, B. L.: 'On the Method of Saddle
      Points', *Applied Scientific Research*, 1950, **B2**, p. 33.
(7) MILLER, W. L., and GORDON, A. R.: 'On the Integral $\int \varepsilon^{z^2} d\mu$
      and Related Integrals', *Journal of Physical Chemistry*,
      1931, **35**, p. 2862.
(8) JAHNKE, E., and EMDE, F.: 'Tables of Functions with
      Formulae and Curves' (Dover Publications, New York,
      1954).
(9) CLEMMOW, P. C., and MUMFORD, C. M.: 'A Table of
      $\sqrt{\left(\dfrac{\pi}{2}\right)} \varepsilon^{\frac{1}{2} j\pi\rho^2} \int_\rho^\infty \varepsilon^{-j\frac{1}{2}\pi\lambda^2} d\lambda$ for Complex Values of ρ',
      Philosophical Transactions of the Royal Society, 1952,
      245, p. 189.
(10) ROSSER, J. B.: 'Theory and Application of $\int_0^z \varepsilon^{-x^2} dx$ and
      $\int_0^z \varepsilon^{-p^2 y^2} dy \int_0^y \varepsilon^{-x^2} dx$' (Mapleton House, New York,
      1948).
(11) GAJEWSKI, R.: 'On Transients in Waveguides', *Bulletin
      de l'Académie Polonaise des Sciences*, 1955, **3**, No. 1, p. 29.

AN ANALYSIS OF TOOTH-RIPPLE PHENOMENA IN SMOOTH LAMINATED POLE-SHOES

By Prof. H. BONDI, M.A., and K. C. MUKHERJI, B.E., Ph.D., Graduate.

SUMMARY

Classical linear electromagnetic theory is used to investigate the mode of penetration of tooth-ripple flux into smooth laminated pole-shoes, without making any assumption as to the thickness of the laminations employed. It is shown that, apart from the mode corresponding to the classical 'depth of penetration' concept, there also exists a completely different mode which penetrates to a much larger radial depth but has a non-uniform distribution across the thickness of the laminations, the maximum values occurring at the interfaces. With thicknesses usually employed in practice, at least 60% of tooth-ripple eddy-current loss is found to arise from this latter mode of penetration, so that only a fraction of the total loss could properly be called a 'pole-face' loss.

(1) LIST OF PRINCIPAL SYMBOLS

The rationalized M.K.S. system of units is used throughout the paper.

B_t = Tooth-ripple flux density at the pole face.
d = Classical 'depth of penetration'.
ε = Base of natural logarithms.
h = Thickness of lamination.
H = Magnetic intensity.
$\bar{H}$ = Complex conjugate of H.
$j = \sqrt{-1}$.
J = Current density.
λ = Slot-pitch.
$k = 2\pi/\lambda$.
P_e = Eddy-current loss per unit pole-face area.
s = Surface area.
v = Peripheral velocity.
V = Volume enclosed by surface s.
μ = Absolute permeability.
μ_0 = Permeability of free space.
μ_r = Relative permeability.
ρ = Resistivity.
ω = Angular frequency.
Ω = Scalar potential.
∇^2 = Laplacian operator.
$[f(x)]_{x=h/2}$ = Value of $f(x)$ for $x = h/2$.
$(\bar{H} \times J)_x$ = x-component of $(\bar{H} \times J)$.
$[f(x, y, z, t)]$ = Real part of $f(x, y, z, t)$.

$A_1, A_2, a_n, b_m, D, f, K_1, K_2, K_3, K_4$ and q_m are arbitrary constants.
The symbols $\hat{B}_t, g_m, l_m, M, N, U, \alpha_n, \beta_m, \gamma_m, \delta_{m0}, H^{(a)}$, etc., are explained in the text.
Bold-face symbols indicate vector quantities.

(2) INTRODUCTION

Tooth ripples are pulsations superimposed on the main flux waves of all types of dynamo-electric machines in which at least one member is dentated. The magnetic flux density at points on a smooth pole-face opposite a tooth is greater than at those opposite a slot, and, as the slotted armature rotates, these waves of flux travel across the pole face. Each point on the pole face is thus subjected to a pulsating flux density and therefore eddy-current and hysteresis losses occur in the pole shoe; at the normal tooth-ripple frequencies, however, it is the eddy-current loss that by far predominates. The present paper is an attempt at a complete mathematical analysis, for a slightly idealized model machine, of the tooth-ripple flux pulsations in smooth laminated pole-shoes and of the corresponding eddy-current losses, under strictly 'linear' electromagnetic conditions.

The important contributions of F. W. Carter,[2] Spooner and Kinnard,[3] Gibbs[4] and Walker[5] in the field of tooth-ripple 'surface loss' were referred to by Prof. G. W. Carter[6] in a recent paper. However, a failure to take full account of the appropriate boundary conditions of the problem invalidates even Prof. Carter's own results, as has since been pointed out by one of the present authors.[7] It appears from the following analysis that the penetration of the pulsating magnetic field is a rather more complicated phenomenon than was hitherto believed.

(3) ASSUMPTIONS

In order to render the problem mathematically tractable, some simplifications have been introduced (apart from the restriction to 'linear' electromagnetic conditions) which make the model machine differ slightly from an actual one:

(a) The pole shoe is assumed to be built up to an infinite length in the axial direction; this is equivalent to ignoring 'end effects', the 'ends' referring to those of the pole shoe and not to those of individual laminations. This is a justifiable procedure, since, in practice, the number of laminations in a pole shoe is large.

(b) The curvature of the pole face is ignored; it can be shown that, so far as tooth-ripple phenomena are concerned, the curvature is of little practical significance.

(c) The smooth pole-face is taken as a complete cylindrical structure; the analysis, however, should be applicable to a salient pole, provided that the pole arc is large compared with the armature slot-pitch.

Two further assumptions are made which reduce the length of the analysis, but these could be dropped without introducing any essential difficulty. These assumptions are:

(d) The interlaminar gap is taken as a perfect insulator of negligible axial width, preventing any flow of current between adjoining laminations.

(e) Attention is primarily focused on the fundamental peripheral harmonic in the Fourier expansion of the tooth-ripple flux wave. Since the equations encountered are all linear until the dissipation is calculated, they apply equally to the higher peripheral harmonics as to the fundamental; however, in calculating the losses, these harmonics of higher order can be disregarded, at least to a first approximation.

(4) THE FIELD EQUATIONS

The rectangular Cartesian co-ordinates used throughout the paper are defined as follows: X is parallel to the axial, Y to the peripheral and Z to the radial direction. The plane pole-face is taken as the surface $z = 0$; the interior of the pole shoe, denoted

Correspondence on Monographs is invited for consideration with a view to publication.
Prof. Bondi is Professor of Mathematics, King's College, University of London.
Dr. Mukherji was formerly in the Electrical Engineering Department, King's College, University of London, and is now with the Metropolitan-Vickers Electrical Co., Ltd. The paper is based on Report Ref. Z/T106 of the British Electrical and Allied Industries Research Association.

[349]

by $z > 0$, is a medium of constant permeability μ (relative permeability μ_r) and, within each lamination of constant resistivity ρ. The plane $x = 0$ is taken to coincide with the central plane of one of the laminations. Finally, the positive direction of y is taken along the direction of motion of the armature teeth relative to the pole face. If v is the relative peripheral velocity and λ is the slot pitch, it is clear that the tooth ripples and all the associated phenomena are functions of $(vt - y)$; these are also periodic functions of y, at any given time, with period λ. Accordingly, the fundamental peripheral harmonic varies as

$$\mathcal{R}\left[\varepsilon^{j(\omega t - ky)}\right]$$

where
$$k = \frac{2\pi}{\lambda}$$

and
$$\omega = vk$$

For the higher peripheral harmonics, integral multiples of the same exponent occur.

Since in all practical cases the tooth-ripple frequency is low enough for displacement currents to be neglected, the relevant macroscopic field equations of Maxwell, applicable for a quasi-stationary state obtaining in 'linear' isotropic media, can be written as follows:

(a) In the region $z \geqslant 0$,

$$\operatorname{div} H = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\operatorname{curl} H = J \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\operatorname{curl} \rho J = -\mu \frac{\partial H}{\partial t} \quad . \quad . \quad . \quad . \quad (3)$$

and therefore it can be shown that

$$\nabla^2 H = \frac{j}{d^2} H \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and
$$\nabla^2 J = \frac{j}{d^2} J \quad . \quad . \quad . \quad . \quad . \quad (4a)$$

where
$$d = \left(\frac{\rho}{\mu\omega}\right)^{1/2} \quad . \quad . \quad . \quad . \quad (5)$$

(b) In the region $z < 0$,
$$\operatorname{div} H = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$\operatorname{curl} H = 0 \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and therefore writing $H = \operatorname{grad} \Omega \quad . \quad . \quad . \quad . \quad . \quad (8)$

it is easy to show that $\nabla^2 \Omega = 0 \quad . \quad . \quad . \quad . \quad . \quad (9)$

(5) TWO EXTREME CASES

Before considering in detail the practical case of laminations of finite thickness, it is instructive to consider briefly the two extreme cases represented by (a) a solid pole-shoe and (b) a hypothetical laminated pole-shoe in which the laminations employed are 'infinitely thin'.

(5.1) Solid Pole-Shoe

By virtue of the assumption of infinite axial length of the pole shoe, the field in this case is independent of x. Remembering that H must vanish as $z \to +\infty$, the relevant solutions of the field equations are as follows:

For $z \geqslant 0$, $H = A_1 \varepsilon^{-z\left(k^2 + \frac{j}{d^2}\right)^{1/2} + j(\omega t - ky)}$

For $z < 0$, $\Omega = (K_1 \varepsilon^{-kz} + K_2 \varepsilon^{kz}) \varepsilon^{j(\omega t - ky)}$

where A_1 is a vector constant, and K_1 and K_2 are constants, with K_1 representing the pole-face potential due to the inducing field. The components of the vector A_1 can easily be evaluated in terms of K_1^* by applying the well-known boundary conditions obtaining on $z = 0$, together with eqn. (1). It is easy to see that the only component of current density not vanishing is J i.e. the current flow is entirely axial. But the important point to notice is that the 'depth of penetration'† of the field into the medium is equal to d, since in practice k is small compared with $1/d$.

(5.2) 'Infinitely Thin' Laminations

At the other extreme, there is the hypothetical case in which the effect of the laminations is taken to be so strong that the flow of current normal to their interfaces is wholly inhibited, i.e. J_z vanishes everywhere. It will be clear that, by virtue of the continued assumption of zero ratio of interlaminar gap to lamination thickness, a pole shoe employing 'infinitely thin' laminations becomes as homogeneous as a solid pole-shoe, so that the field is again independent of x, as stated in Section 5.1. In spite of eqns. (1) and (6), therefore, H_x is in no way coupled to H_y and H_z; also, eqns. (4), (8) and (9) involve all the three components of H, but only individually. Since the inducing field is independent of x, H_x must vanish as $z \to -\infty$; also, it must vanish as $z \to +\infty$. It follows, therefore, that H_x vanishes throughout in both the zones $z \geqslant 0$ and $z < 0$. It is easy to see then that J_y and J_z also vanish everywhere.

This idealization therefore leads in effect to a medium of infinite volume resistivity, for there occurs a total inhibition of induced currents in the region $z \geqslant 0$ just as is the case in the region $z < 0$. Remembering again that H tends to zero as $z \to +\infty$, the relevant solutions of the field equations in this case are as follows:

For $z \geqslant 0$, $H_x = 0$

$$H_y = jA_2 \varepsilon^{-kz + j(\omega t - ky)}$$

$$H_z = A_2 \varepsilon^{-kz + j(\omega t - ky)}$$

For $z < 0$, $\Omega = (K_3 \varepsilon^{-kz} + K_4 \varepsilon^{kz}) \varepsilon^{j(\omega t - ky)}$

where A_2, K_3 and K_4 are constants, K_3 representing the pole-face potential due to the inducing field. It is again easy to evaluate A_2 in terms of K_3.‡ Quite naturally, under the above hypothetical conditions, the solutions for the quasi-stationary state are exactly the same—except for the factor $\varepsilon^{j\omega t}$—as can readily be obtained for the static (i.e. with the armature at standstill) distribution in any case. It is important, however, to observe that the total inhibition of currents resulting from the use of 'infinitely thin' laminations leads to a quasi-static potential magnetic field penetrating to a 'depth' $1/k$ into the medium, and, in practice, $1/k$ is very large compared with the classical 'depth' of penetration', d.

The essential phenomena underlying the classical 'depth of penetration' concept are that a varying magnetic field induces currents in the surface layers of a conducting medium subjected to it, such that these currents give rise to an opposing magnetic field and thus shield the interior of the conductor from the inducing field. However, if it were possible to prevent such currents from flowing, by some hypothetic means like employ-

---

* Ignoring the demagnetizing effect of the induced eddy currents, K_1 in turn can be shown to be related to the amplitude $\hat{B}_t$ of the static tooth-ripple flux density at the pole face by the equation

$$K_1 = -\left[\frac{k\mu_r + \left(k^2 + \frac{j}{d^2}\right)^{1/2}}{2k^2\mu}\right]\hat{B}_t$$

† In the paper, the 'depth of penetration' is defined as the depth which would contain the entire flux if the magnetic intensity throughout this depth were uniform and equal to the intensity at the surface of the lamination.

‡ It can be shown that $K_3 = -\left(\frac{\mu_r + 1}{2k\mu}\right)\hat{B}_t$.

g 'infinitely thin' laminations or 'permeable insulators', the
shielding effect would be absent and the field would therefore
penetrate to a much larger 'depth', i.e. it would suffer a far less
sharp attenuation, within the medium.

(6) THE GENERAL CASE OF LAMINATIONS OF FINITE THICKNESS

In practice, laminated pole-shoes employ laminations of some
finite thickness. However, the deliberate result is still a partial
prevention of the flow of induced currents, leading to a corre-
sponding reduction of their shielding effect. A deeper penetra-
tion of the pulsations can therefore be expected to occur in a
laminated pole-shoe than in a solid one. In the practical case
that is discussed below, there are shown to be simultaneously
present in a practical laminated pole-shoe both of the con-
trasting modes of field penetration—one attenuating much more
sharply than the other—as envisaged in Sections 5.1 and 5.2,
respectively.

(6.1) Formulation of Boundary Conditions: Proximity Effect

In the general case, all the field variables depend on x; that
this should be so is due entirely to the laminated construction of
the pole shoe. As the laminations repeat with thickness h and
any effects due to the open ends of the pole-shoe are ignored, it
follows that all the field variables are—both inside and outside
the pole shoe—periodic functions of x, with period h. More-
over, the field variables have to satisfy certain symmetry con-
ditions in their dependence on x, which arise as follows:

(6.1.1) Zone $-\frac{h}{2} \leqslant x \leqslant +\frac{h}{2}$; region $z \geqslant 0$.

H_y and H_z would be independent of x in the absence of the
laminations; their dependence on x arises solely from the
insulating property of the interfaces at $x = \pm h/2$. Since this
feature is symmetrically placed round $x = 0$, it follows that
the changes in H_y and H_z are similarly symmetrical, so that H_y
and H_z are even functions of x. This implies that

$$[H_y]_{x=h/2} = [H_y]_{x=-(h/2)}$$

and the periodicity implies that H_y is continuous at the interfaces
$x = \pm h/2$; the same holds for H_z also. The derivatives $\partial H_y/\partial x$
and $\partial H_z/\partial x$, however, need not be continuous at the interfaces,
since a discontinuity there in these derivatives merely implies
that the current densities J_y and J_z differ on different sides of
the interfaces, which is an entirely plausible situation.

Since H_y and H_z are even functions of x, as established above,
J_x then must be an odd function of x, by eqn. (1). This implies
that

$$[H_x]_{x=h/2} = -[H_x]_{x=-(h/2)}$$

and since there can be no discontinuity at the interfaces in the
component of H normal to them, it follows from the repetition
of H in consecutive laminations that H_x must vanish on
$x = \pm h/2$ (Fig. 1). These symmetry properties of the com-
ponents of H imply, by eqn. (2), that J_y and J_z are odd functions
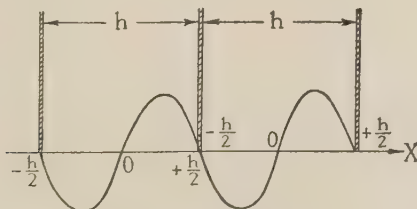
Fig. 1.—Variation of H_x, $m = 1$.
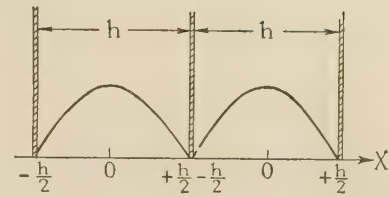
Fig. 2.—Variation of J_x, $n = 0$.

of x while J_x is an even function. It is clear also that J_x must
vanish at $x = \pm h/2$ (Fig. 2), since the interfaces are taken to
be non-conducting.

(6.1.2) Region $z < 0$: the Air-Gap.

In this zone there are no interfaces at $x = \pm h/2$, and per-
fectly ordinary conditions prevail there. Accordingly H, while
obeying the periodicity and symmetry conditions imposed by
the laminations in the zone $z \geqslant 0$, must be continuous at
$x = \pm h/2$, with continuous spatial derivatives of all orders.
Also, in the absence of the laminations, H_x would vanish, since
it is entirely due to the currents induced in the pole shoe.
Accordingly in the air-gap far from the pole face (this is mathe-
matically expressed by $z \to -\infty$), H_x must tend to zero. The
same argument applies to the derivatives $\partial H_y/\partial x$ and $\partial H_z/\partial x$,
since, but for the laminations, the field would be independent
of x. The above considerations take full account of the proximity
effect of adjoining laminations.

Finally, the boundary conditions on $z = 0$ remain to be
considered. The tangential components of H and the normal
components both of J and of the induction vector μH must be
continuous on this boundary. Hence

$$[J_z]_{z=0} = 0 \quad \cdots \quad \cdots \quad (10)$$

$$[H_x]_{z\to 0+0} = [H_x]_{z\to 0-0} \quad \cdots \quad (11)$$

$$[H_y]_{z\to 0+0} = [H_y]_{z\to 0-0} \quad \cdots \quad (12)$$

$$[\mu H_z]_{z\to 0+0} = [\mu_0 H_z]_{z\to 0-0} \quad \cdots \quad (13)$$

And, as $z \to +\infty$, i.e. far from the pole face inside the pole shoe,
all the field variables tend to zero $\quad \cdots \quad \cdots \quad (14)$

(6.2) Solution of Field Equations

(6.2.1) Region $z \geqslant 0$: the Pole Shoe.

Inside the pole shoe, the following three types of fields can be
distinguished:

(a) $H_x = 0$ for all values of x, but J_x is not zero.

(b) $J_x = 0$ for all values of x, but H_x is not zero.

(c) $H_x = 0$ and $J_x = 0$, for all values of x.*

In fact, by the well-known principle of superposition of fields
on a 'linear' electromagnetic circuit, *any* general tooth-ripple
field can be represented as a sum of these three types of field.

(6.2.1.1) Field of Type (a).

The restrictions on J_x are, as stated in Section 6.1, that J_x
must be a periodic even function of x with period h, and it must
also vanish on $x = \pm h/2$. Instead of expanding J_x in a Fourier
series in x of period h, it is more convenient to expand it in a
Fourier series of period $2h$, although, of course, this series
represents J_x only in the range $-h/2 \leqslant x \leqslant +h/2$ and not
beyond it. For such a series, the condition that J_x must vanish

* The field of type (c) is evidently included within types (a) and (b); it is singled out
here as a separate type since it becomes of special importance, as will be seen below.

on $x = \pm h/2$ simply implies that only the odd-order terms remain, namely terms of the type

$$\cos\left[(2n+1)\frac{\pi x}{h}\right]$$

n being any integer including zero. It follows from eqn. (4) and condition (14) that the dependence of such a term on x, y, z and t is of the form

$$\cos\left[(2n+1)\frac{\pi x}{h}\right]\varepsilon^{-z\left[k^2+(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}\right]^{1/2}+j(\omega t-ky)}$$

Adding all such terms, each multiplied by an arbitrary constant a_n, it is established that

$$J_x = \sum_{n=0}^{\infty} a_n \cos\left[(2n+1)\frac{\pi x}{h}\right]\varepsilon^{-z\left[k^2+(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}\right]^{1/2}+j(\omega t-ky)}$$

Since $H_x \equiv 0$ for fields of this type, it follows from eqn. (1) that

$$\frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} = 0$$

also, by eqn. (2),

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = J_x$$

These equations may readily be solved, and the solution not having the above type of dependence on x may be omitted

— H_y, H_z, J_x
--- J_y, J_z
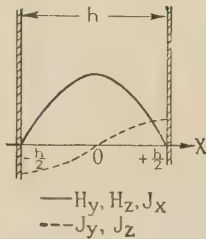
Fig. 3.—Field of type (a): $H_x = 0$, $n = 0$.

since it will be included in the field of type (c). These results (see Fig. 3) are as follows:

$$H_x^{(a)} = 0$$

$$H_y^{(a)} = \sum_{n=0}^{\infty} \frac{a_n\left[k^2+(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}\right]^{1/2}}{(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}} \times$$

$$\times \cos\left[(2n+1)\frac{\pi x}{h}\right]\varepsilon^{-z\left[k^2+(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}\right]^{1/2}+j(\omega t-ky)}$$

$$H_z^{(a)} = -jk\sum_{n=0}^{\infty} \frac{a_n}{(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}}$$

$$\times \cos\left[(2n+1)\frac{\pi x}{h}\right]\varepsilon^{-z\left[k^2+(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}\right]^{1/2}+j(\omega t-ky)}$$

$$J_x^{(a)} = \sum_{n=0}^{\infty} a_n \cos\left[(2n+1)\frac{\pi x}{h}\right]\varepsilon^{-z\left[k^2+(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}\right]^{1/2}+j(\omega t-ky)}$$

$$J_y^{(a)} = -j\frac{\pi k}{h}\sum_{n=0}^{\infty} \frac{a_n(2n+1)}{(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}}$$

$$\times \sin\left[(2n+1)\frac{\pi x}{h}\right]\varepsilon^{-z\left[k^2+(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}\right]^{1/2}+j(\omega t-ky)}$$

$$J_z^{(a)} = -\frac{\pi}{h}\sum_{n=0}^{\infty} \frac{a_n(2n+1)\left[k^2+(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}\right]^{1/2}}{(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}}$$

$$\times \sin\left[(2n+1)\frac{\pi x}{h}\right]\varepsilon^{-z\left[k^2+(2n+1)^2\frac{\pi^2}{h^2}+\frac{j}{d^2}\right]^{1/2}+j(\omega t-k}$$

(6.2.1.2) *Field of Type (b).*

In this type of field, H_x is to be a periodic odd function of x of period h, and is to vanish at $x = \pm h/2$. Hence H_x can be expanded in a Fourier series with terms of the type

$$\sin\left(2m\frac{\pi x}{h}\right)$$

m being any integer excluding zero. Eqn. (4) and condition (14) determine such terms as being proportional to

$$\varepsilon^{-z\left[k^2+\left(\frac{2m\pi}{h}\right)^2+\frac{j}{d^2}\right]^{1/2}}$$

Multiplying each term by an arbitrary constant b_m and solving

— H_x, J_y, J_z
--- H_y, H_z
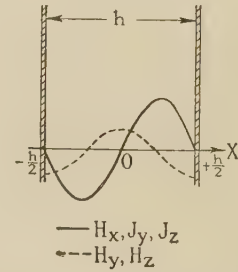
Fig. 4.—Field of type (b): $J_x = 0$, $m = 1$.

for H_y and H_z from eqns. (1) and (2), and again omitting solutions with any different type of dependence on x, the results (see Fig. 4) are as follows:

$$H_x^{(b)} = \sum_{m=1}^{\infty} b_m \sin\left(2m\frac{\pi x}{h}\right)\varepsilon^{-z\left[k^2+\left(2m\frac{\pi}{h}\right)^2+\frac{j}{d^2}\right]^{1/2}+j(\omega t-ky)}$$

$$H_y^{(b)} = j\frac{2\pi k}{h}\sum_{m=1}^{\infty} \frac{b_m m}{\left(2m\frac{\pi}{h}\right)^2+\frac{j}{d^2}}$$

$$\times \cos\left(2m\frac{\pi x}{h}\right)\varepsilon^{-z\left[k^2+\left(2m\frac{\pi}{h}\right)^2+\frac{j}{d^2}\right]^{1/2}+j(\omega t-ky}$$

$$H_z^{(b)} = \frac{2\pi}{h}\sum_{m=1}^{\infty} \frac{b_m m\left[k^2+\left(2m\frac{\pi}{h}\right)^2+\frac{j}{d^2}\right]^{1/2}}{\left(2m\frac{\pi}{h}\right)^2+\frac{j}{d^2}}$$

$$\times \cos\left(2m\frac{\pi x}{h}\right)\varepsilon^{-z\left[k^2+\left(2m\frac{h}{\pi}\right)^2+\frac{j}{d^2}\right]^{1/2}+j(\omega t-ky}$$

$$J_x^{(b)} = 0$$

$$J_y^{(b)} = -\frac{j}{d^2}\sum_{m=1}^{\infty} \frac{b_m\left[k^2+\left(2m\frac{\pi}{h}\right)^2+\frac{j}{d^2}\right]^{1/2}}{\left(2m\frac{\pi}{h}\right)^2+\frac{j}{d^2}}$$

$$\times \sin\left(2m\frac{\pi x}{h}\right)\varepsilon^{-z\left[k^2+\left(2m\frac{\pi}{h}\right)^2+\frac{j}{d^2}\right]^{1/2}+j(\omega t-ky}$$

$$b) = -\frac{k}{d^2} \sum_{m=1}^{\infty} \frac{b_m}{\left(2m\frac{\pi}{h}\right)^2 + \frac{j}{d^2}}$$

$$\times \sin\left(2m\frac{\pi x}{h}\right) \varepsilon^{-z\left[k^2 + \left(2m\frac{\pi}{h}\right)^2 + \frac{j}{d^2}\right]^{1/2} + j(\omega t - ky)}$$

..2.1.3) *Field of Type (c).*

Since, in this type of field,

$$H_x = 0 \quad \text{and} \quad J_x = 0$$

follows from eqns. (1) and (2) that

$$\frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} = 0$$

ad

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = 0$$

ence if U is a function, also proportional to

$$\mathcal{R}\left[\varepsilon^{j(\omega t - ky)}\right]$$

ich that

$$H_y = \frac{\partial U}{\partial y}$$

follows that

$$H_z = \frac{\partial U}{\partial z}$$

na

$$\frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = 0$$

herefore

$$U \propto \varepsilon^{-kz + j(\omega t - ky)}$$

lso, $U(x)$ satisfies an equation of the form of eqn. (4), namely

$$\frac{\partial^2 U(x)}{\partial x^2} = \frac{j}{d^2} U(x)$$
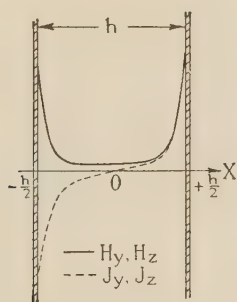
Fig. 5.—Field of type (c): $J_x = 0$, $H_x = 0$.

Since H_y and H_z, and therefore U, are even functions of x, follows that

$$U \propto \cosh\left[(1 + j)\frac{x}{\sqrt{(2)}d}\right]$$

n the range $-\frac{h}{2} \leqslant x \leqslant +\frac{h}{2}$.

With an arbitrary constant f, the results (see Fig. 5) are as bllows:

$$H_x^{(c)} = 0$$

$$H_y^{(c)} = -jkf \cosh\left[(1 + j)\frac{x}{\sqrt{(2)}d}\right]\varepsilon^{-kz + j(\omega t - ky)}$$

$$H_z^{(c)} = -kf \cosh\left[(1 + j)\frac{x}{\sqrt{(2)}d}\right]\varepsilon^{-kz + j(\omega t - ky)}$$

$$J_x^{(c)} = 0$$

$$J_y^{(c)} = (1 + j)\frac{kf}{\sqrt{(2)}d} \sinh\left[(1 + j)\frac{x}{\sqrt{(2)}d}\right]\varepsilon^{-kz + j(\omega t - ky)}$$

$$J_z^{(c)} = (1 - j)\frac{kf}{\sqrt{(2)}d} \sinh\left[(1 + j)\frac{x}{\sqrt{(2)}d}\right]\varepsilon^{-kz + j(\omega t - ky)}$$

It will be observed that, whereas the fields of types (a) and (b) attenuate in the radial direction at least as rapidly as indicated by the conventional 'depth of penetration' concept, the field of type (c) behaves quite differently. It penetrates to a much greater depth $1/k$ in the radial direction but falls off along the axial direction on either side of the lamination interfaces; this type of field is essentially spread along the interfaces for a sufficiently thick lamination, where this 'skin effect' is more prominent.

All the above contributions to the component vectors from each type of field satisfy individually eqns. (1)–(4a) and also condition (14), the contributions to J_x and H_x duly vanishing at the interfaces. The complete field-intensity and current-density vector components are obtained as the sum of the respective contributions from each type of field.

(6.2.2) **Region $z < 0$: the Air-Gap.**

The air-gap field potential Ω must be a periodic even function of x, of period h, since H and all its spatial derivatives have to be continuous at the planes $x = \pm h/2$ in the air gap. Accordingly, from eqn. (9),

$$\Omega = \sum_{m=0}^{\infty} q_m \cos\left(2m\frac{\pi x}{h}\right) \varepsilon^{z\left[k^2 + \left(2m\frac{\pi}{h}\right)^2\right]^{1/2} + j(\omega t - ky)}$$
$$+ D \sinh(kz)\varepsilon^{j(\omega t - ky)}$$

where D and q_m are arbitrary constants.* The positive sign of the coefficient of z in the above series is determined by the condition that H_x, i.e. $\partial\Omega/\partial x$, must vanish as $z \to -\infty$ (see Section 6.1), and the last term in the expression for Ω arises because this same condition does not affect any part of Ω which is independent of x. In fact, the ε^{-kz} part of this last term represents the inducing potential field that produces the entire phenomenon, since all the other field components in the air-gap diminish with distance from the pole face; the chief task is then to determine a_n, b_m, q_m and f in terms of D. The coefficient D can in turn be expressed in terms of the static tooth-ripple flux density at the pole face, as shown in eqn. (17).

The boundary condition [eqn. (12)] is next satisfied, considering the sum of fields (a), (b) and (c) inside the pole shoe and the Ω-field outside. Solving the resulting equation for q_m, it can be established[8] that

$$\Omega = \varepsilon^{j(\omega t - ky)} \sum_{m=0}^{\infty} (-1)^m \cos\left(2m\frac{\pi}{h}\right)\varepsilon^{z\left[k^2 + \left(2m\frac{\pi}{h}\right)^2\right]^{1/2}}$$

$$\left\{ j\frac{4}{\pi k}(1 - \tfrac{1}{2}\delta_{m0}) \sum_{n=0}^{\infty} (-1)^n \frac{a_n(2n + 1)}{(2n + 1)^2 - 4m^2} \right.$$

$$\times \frac{\left[k^2 + (2n + 1)^2\frac{\pi^2}{h^2} + \frac{j}{d^2}\right]^{1/2}}{(2n + 1)^2\frac{\pi^2}{h^2} + \frac{j}{d^2}} + (-1)^{m+1}\frac{b_m\frac{2m\pi}{h}}{\left(2m\frac{\pi}{h}\right)^2 + \frac{j}{d^2}}$$

$$\left. + (1 - \tfrac{1}{2}\delta_{m0})(1 + j)\frac{2\sqrt{2}}{dh}f\frac{\sinh\left[(1 + j)\frac{h}{2\sqrt{(2)}d}\right]}{\left(2m\frac{\pi}{h}\right)^2 + \frac{j}{d^2}} \right\}$$

$$+ D \sinh(kz)\varepsilon^{j(\omega t - ky)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

* All intermediate algebraic steps have been omitted from the paper owing to lack of space; these details will be found in Reference 8.

where $$b_0 = 0$$

and
$$\delta_{m0} = \begin{cases} 1 \text{ for } m = 0 \\ 0 \text{ for } m \neq 0 \end{cases}$$

H_x, H_y and H_z are obtained from eqn. (15) by virtue of eqn. (8). The undamped tooth-ripple flux density B_t at the pole face is given by*

$$B_t = \mu_0 \left[\frac{\partial}{\partial z} [D \sinh (kz) \varepsilon^{j(\omega t - ky)}] \right]_{z=0}$$

$$= \mu_0 k D \varepsilon^{j(\omega t - ky)} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

so that
$$D = \frac{\hat{B}_t}{\mu_0 k} \quad . \quad . \quad . \quad . \quad . \quad (17)$$

$\hat{B}_t$ denoting the amplitude of B_t, and therefore also the amplitude of the static tooth-ripple flux density at the pole face.

(6.3) Complete Expressions for Vector Components for the Region $z \geqslant 0$

The evaluation of the arbitrary constants a_n, b_m and f in terms of $\hat{B}_t$ is performed by making use of eqns. (10) and (13). The algebra involved is rather complicated, even after considerable simplification has been made by imposing certain restrictions on the values of the parameters (but only such as are quite legitimate for laminations usually employed in practice, namely μ_r of the order of 100, h/d of the order of 10 and kh of the order of unity); the essentials of the work will be found elsewhere.[8] Expressions are then obtained for the current-density and magnetic-intensity vector components. Using the definitions

$$g_m \left(\frac{\pi d}{h} \right) = \sum_{n=0}^{\infty} \frac{\left[j + (2n+1)^2 \frac{\pi^2 d^2}{h^2} \right]^{-3/2}}{(2n+1)^2 - 4m^2}$$

$$l_m \left(\frac{\pi d}{h} \right) = \sum_{n=0}^{\infty} \frac{\left[j + (2n+1)^2 \frac{\pi^2 d^2}{h^2} \right]^{-1/2}}{(2n+1)^2 [(2n+1)^2 - 4m^2]}$$

$$\alpha_n = \left(\left\{ [\sqrt{(2)} - kd] + jkd \right\} \cot \left[(1-j) \frac{h}{2\sqrt{(2)}d} \right] \right.$$
$$\left. + \frac{2\sqrt{(2)}kh}{\pi^2} \frac{\left[j + (2n+1)^2 \frac{\pi^2 d^2}{h^2} \right]}{(2n+1)^2} \right)$$

$$\beta_m = \left(\left[j + \left(2m \frac{\pi d}{h} \right)^2 \right]^{-1} + (1-j) \frac{2kh}{\pi^2} \times \right.$$
$$\times \left\{ [\sqrt{(2)} - kd] + jkd \right\} \cot \left[(1-j) \frac{h}{2\sqrt{(2)}d} \right] g_m \left(\frac{\pi d}{h} \right)$$
$$\left. + (1-j) \frac{4\sqrt{(2)}k^2 h^2}{\pi^4} l_m \left(\frac{\pi d}{h} \right) \right)$$

$$\gamma_m = \left\{ \beta_m - (1+j) \frac{h}{2\sqrt{(2)}\pi \mu_r d} \frac{1}{m} \left[j + \left(2m \frac{\pi d}{h} \right)^2 \right]^{-1} \right\}$$

* Strictly, the term independent of x (corresponding to $m = 0$) in the series in eqn. (15) also contributes towards $\hat{B}_t$, leading to the exact result:
$$D = \left(1 + \frac{1}{\mu_r} \right) \frac{\hat{B}_t}{\mu_0 k}$$
Since μ_r is quite large in practice, the error involved in eqn. (17) is negligible.

these components, for the region $z \geqslant 0$, can be written dow as follows:

$$J_x = J_x^{(a)}$$
$$= (1+j) \frac{h}{\pi d^2} \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \sum_{n=0}^{\infty} (-1)^n$$
$$\times \frac{\alpha_n \left[j + (2n+1)^2 \frac{\pi^2 d^2}{h^2} \right]^{-1/2}}{2n+1} \cos \left[(2n+1) \frac{\pi x}{h} \right]$$
$$\times \varepsilon^{-z \left[k^2 + (2n+1)^2 \frac{\pi^2}{h^2} + \frac{j}{d^2} \right]^{1/2}} \quad . \quad (18)$$

$$J_y = J_y^{(a)} + J_y^{(b)} + J_y^{(c)}$$
$$= (1-j)k \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \sum_{n=0}^{\infty} (-1)^n \alpha_n \left[j + (2n+1)^2 \frac{\pi^2 d^2}{h^2} \right]^{-3/2}$$
$$\times \sin \left[(2n+1) \frac{\pi x}{h} \right] \varepsilon^{-z \left[k^2 + (2n+1)^2 \frac{\pi^2}{h^2} + \frac{j}{d^2} \right]^{1/2}}$$
$$+ j \frac{h}{\pi d^2} \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \sum_{m=0}^{\infty} (-1)^m \frac{\gamma_m}{m} \sin \left(2m \frac{\pi x}{h} \right)$$
$$\times \varepsilon^{-z \left[k^2 + \left(2m \frac{\pi}{h} \right)^2 + \frac{j}{d^2} \right]^{1/2}} - \frac{h}{2d^2} \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)}$$
$$\times \frac{\sinh \left[(1+j) \frac{x}{\sqrt{(2)}d} \right]}{\sinh \left[(1+j) \frac{h}{2\sqrt{(2)}d} \right]} \varepsilon^{-kz} \quad . \quad . \quad . \quad (19)$$

$$J_z = J_z^{(a)} + J_z^{(b)} + J_z^{(c)}$$
$$= (1+j) \frac{1}{d} \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \sum_{n=0}^{\infty} (-1)^{n+1} \alpha_n$$
$$\times \left[j + (2n+1)^2 \frac{\pi^2 d^2}{h^2} \right]^{-1} \sin \left[(2n+1) \frac{\pi x}{h} \right]$$
$$\times \varepsilon^{-z \left[k^2 + (2n+1)^2 \frac{\pi^2}{h^2} + \frac{j}{d^2} \right]^{1/2}} + \frac{kh}{\pi d} \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)}$$
$$\times \sum_{m=1}^{\infty} (-1)^m \frac{\gamma_m}{m} \left[j + \left(2m \frac{\pi d}{h} \right)^2 \right]^{-1/2} \sin \left(2m \frac{\pi x}{h} \right)$$
$$\times \varepsilon^{-z \left[k^2 + \left(2m \frac{\pi}{h} \right)^2 + \frac{j}{d^2} \right]^{1/2}} + j \frac{h}{2d^2} \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)}$$
$$\times \frac{\sinh \left[(1+j) \frac{x}{\sqrt{(2)}d} \right]}{\sinh \left[(1+j) \frac{h}{2\sqrt{(2)}d} \right]} \varepsilon^{-kz} \quad . \quad . \quad . \quad . \quad (20)$$

$$H_x = H_x^{(b)}$$
$$= \frac{h}{\pi d} \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \sum_{m=1}^{\infty} (-1)^{m+1} \frac{\gamma_m}{m} \left[j + \left(2m \frac{\pi d}{h} \right)^2 \right]^{1/2}$$
$$\times \sin \left(2m \frac{\pi x}{h} \right) \varepsilon^{-z \left[k^2 + \left(2m \frac{\pi}{h} \right)^2 + \frac{j}{d^2} \right]^{1/2}} \quad . \quad . \quad . \quad . \quad (21)$$

$$H_y = H_y^{(a)} + H_y^{(b)} + H_y^{(c)}$$
$$= (1+j) \frac{h}{\pi d} \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \sum_{n=0}^{\infty} (-1)^n$$

$$\times \frac{\alpha_n \left[j + (2n+1)^2 \frac{\pi^2 d^2}{h^2} \right]^{-1}}{2n+1} \cos \left[(2n+1)\frac{\pi x}{h} \right]$$

$$\times \varepsilon^{-z\left[k^2 + (2n+1)^2 \frac{\pi^2}{h^2} + \frac{j}{d^2} \right]^{1/2}} + j2kd \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \sum_{m=1}^{\infty} (-1)^{m+1}$$

$$\times \gamma_m \left[j + \left(2m\frac{\pi d}{h} \right)^2 \right]^{-1/2} \cos \left(2m\frac{\pi x}{h} \right) \varepsilon^{-z\left[k^2 + \left(2m\frac{\pi}{h} \right)^2 + \frac{j}{d^2} \right]^{1/2}}$$

$$+ (1+j)\frac{h}{2\sqrt{(2)}d} \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \frac{\cosh \left[(1+j)\frac{x}{\sqrt{(2)}d} \right]}{\sinh \left[(1+j)\frac{h}{2\sqrt{(2)}d} \right]} \varepsilon^{-kz}$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad (22)$$

$$H_z = H_z^{(a)} + H_z^{(b)} + H_z^{(c)}$$

$$= (1-j)\frac{kh}{\pi} \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \sum_{m=0}^{\infty} (-1)^n \frac{\alpha_n}{2n+1}$$

$$\times \left[j + (2n+1)^2 \frac{\pi^2 d^2}{h^2} \right]^{-3/2} \cos \left[(2n+1)\frac{\pi x}{h} \right]$$

$$\times \varepsilon^{-z\left[k^2 + (2n+1)^2 \frac{\pi^2}{h^2} + \frac{j}{d^2} \right]^{1/2}} + 2\frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \sum_{m=0}^{\infty} (-1)^{m+1} \gamma_m$$

$$\times \cos \left(2m\frac{\pi x}{h} \right) \varepsilon^{-z\left[k^2 + \left(2m\frac{\pi}{h} \right)^2 + \frac{j}{d^2} \right]^{1/2}} + (1-j)\frac{h}{2\sqrt{(2)}d}$$

$$\times \frac{\hat{B}_t}{\mu \beta_0} \varepsilon^{j(\omega t - ky)} \frac{\cosh \left[(1+j)\frac{x}{\sqrt{(2)}d} \right]}{\sinh \left[(1+j)\frac{h}{2\sqrt{(2)}d} \right]} \varepsilon^{-kz} \quad \cdot \quad \cdot \quad \cdot \quad (23)$$

The expressions for H_y and H_z in eqns. (22) and (23) indicate that only parts of the tooth-ripple field penetrate a lamination to the order of the conventional radial 'depth' d from the pole face. With a sufficiently thick lamination, a major part of the field penetrates it to a lateral depth of the same order from its interfaces with the adjoining laminations, and its attenuation with radial depth is comparatively much slower (Fig. 6). This is the crucial result of the above analysis.

$\overline{H}$ denoting the complex conjugate of H. Using the symbols

$$M = \mathcal{R}[\overline{H}_y^{(c)} J_z^{(c)} - \overline{H}_z^{(c)} J_y^{(c)} + \overline{H}_y^{(c)} J_z^{(a)} - \overline{H}_z^{(c)} J_y^{(a)}$$
$$+ \overline{H}_y^{(b)} J_z^{(c)} - \overline{H}_z^{(b)} J_y^{(c)} + \overline{H}_y^{(b)} J_z^{(a)} - \overline{H}_z^{(b)} J_y^{(a)}]_{x=h/2}$$

and

$$N = \mathcal{R}[\overline{H}_y J_x^{(a)} - \overline{H}_x^{(b)} J_y]_{z=0}$$

it can be shown[8] that the tooth-ripple eddy-current loss per unit area of pole face is given by

$$P_e = \frac{\rho}{h} \left[\int_0^{\infty} M dz + \int_0^{h/2} N dx \right] \quad \cdot \quad \cdot \quad \cdot \quad (24)$$

(8) RESULTS AND COMMENTS

Although it has not been possible in Section 7 to present a formula for the tooth-ripple eddy-current loss in a directly comprehensible form, in terms of the basic parameters,* the process of loss calculation is not, in fact, as difficult and time consuming as it might appear; indeed it becomes quite simple if the laminations are rather thin. In such cases, the axial component of the magnetic intensity becomes negligible,

i.e. $\qquad b_m \simeq 0$

considerable simplification being thereby achieved,

namely $\qquad \alpha_n \simeq \sqrt{(2)} \cot \left[(1-j)\frac{h}{2\sqrt{(2)}d} \right]$

for all values of n;

$$\beta_0 \simeq \left\{ -j + (1-j)\frac{2\sqrt{(2)}kh}{\pi^2} \cot \left[(1-j)\frac{h}{2\sqrt{(2)}d} \right] g_0 \left(\frac{\pi d}{h} \right) \right\}$$

and $\gamma_m \simeq 0$, for all values of m.

For an experimental machine in which

$\lambda = 2 \cdot 44 \times 10^{-2}$ metre.
$\mu_r = 1\,200$ (assumed).
$\omega = 3\,770$ rad/sec.
$\rho = 20 \times 10^{-8}$ ohm-metre (assumed).

it is found possible to admit of such simplification with practical

Fig. 6.—Tooth-ripple flux penetration in a thick lamination: Lines of constant $\hat{H}$: (Not to scale).

(7) EDDY-CURRENT LOSS

In view of the complicated nature of the expressions obtained for the current-density vector components, it is more convenient to make use of the Poynting vector for the calculation of the eddy-current loss. Thus, in the absence of hysteresis,

$$\text{Eddy-current loss} = -\rho \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \int_s [\mathcal{R}(J \times H) \cdot ds] dt$$

$$= \frac{\rho}{2} \left[\mathcal{R} \int_s (\overline{H} \times J) \cdot ds \right]$$

impunity for a lamination thickness not exceeding $1 \cdot 6$ mm. However, the appropriate resistivity and permeability figures being subject to considerable variation, it is difficult to lay down a rigid rule. It is not possible either to foresee from eqn. (24) the influence on the losses of the various parameters involved, until a sufficiently large number of calculations have been made.

From sample calculations[8] for the above experimental machine with four different thicknesses (namely $0 \cdot 40$, $0 \cdot 80$, $1 \cdot 59$ and $3 \cdot 18$ mm) of pole punchings regarded otherwise as similar, it

* It may be possible to present at a later date curves for certain modified parameters in terms of basic, dimensionless parameters like h/d and kh, which would permit of a much quicker estimation of the loss.

can be confirmed, however, that, even in the thickest punching considered, more than 60% of the total tooth-ripple eddy-current loss arises from the mode of pulsation that penetrates deeply into the pole shoe. The term 'pole-face loss', meaning a loss occurring at the air-gap surface of the pole shoe, would thus appear to be a misnomer if used to indicate the total loss in such laminated pole-shoes.

Also, in the 0·40–0·80 mm range of lamination thickness, the total calculated eddy-current loss is found to be roughly proportional to $h^{1·8}$, while in the 0·80–3·18 mm range the loss varies approximately linearly with thickness (see Fig. 7). This is

Fig.7.—Variation of calculated tooth-ripple eddy-current loss with thickness of lamination.

in close accord with the relevant experimental observations of Spooner and Kinnard[3] and other investigators, which have so far defied a rational quantitative explanation.

(9) CONCLUSION

The 'skin-effect' phenomenon across the thickness of a pole punching associated with tooth-ripple flux pulsation had indeed been anticipated, even if only qualitatively, by Adams and his collaborators[1] and also by Spooner and Kinnard.[3] The contribution to the vector components from the field distinguished above as of type (c) gives a quantitative measure of this effect, ignoring hysteresis. It has indeed been obvious for a long time

that the thickness of the pole punchings is one of the major factors governing the tooth-ripple losses, but the exact character of the part it plays has remained rather obscure. It is believed that, in spite of the limitations of 'linear' macroscopic electromagnetic theory, the present work helps to throw new light on the phenomenon of tooth-ripple eddy-current loss and to explain some of its more obscure aspects. The results of an experimental investigation described in a companion paper[9] tend to corroborate the theory outlined in the present paper, namely, that an appreciable amount of tooth-ripple flux pulsation penetrates deep into the pole shoe and that such penetration is practically independent of the tooth-ripple frequency.

(11) REFERENCES

(1) ADAMS, C. A., LANIER, A. C., POPE, C. C., and SCHOOLEY, C. O.: 'Pole-face Losses', Transactions of the American I.E.E., 1909, 28, p. 1133.
(2) CARTER, F. W.: 'Pole-Face Losses', Journal I.E.E., 1916, 54, p. 168.
(3) SPOONER, T., and KINNARD, I. F.: 'Surface Iron Losses with Reference to Laminated Materials', Transactions of the American I.E.E., 1924, 43, p. 262.
(4) GIBBS, W. J.: 'Tooth-Ripple Losses in Unwound Pole Shoes', Journal I.E.E., 1947, 94, Part II, p. 2.
(5) WALKER, J. H.: 'A Theory of Induction Motor Surface Losses', ibid., 1948, 95, Part II, p. 597.
(6) CARTER, G. W.: 'A Note on the Surface Loss in a Laminated Pole-Face', Proceedings I.E.E., Monograph No. 123, March, 1955 (102, C, p. 217).
(7) MUKHERJI, K. C.: Discussion on Reference 6, ibid., 1956, 103, Part C, p. 230.
(8) BONDI, H., and MUKHERJI, K. C.: 'An Analysis of Tooth-ripple Phenomena in Smooth Laminated Pole-shoes', E.R.A. Technical Report Z/T106, 1956.
(9) GREIG, J., and MUKHERJI, K. C.: 'An Experimental Investigation of Tooth-ripple Flux Pulsations in Smooth Laminated Pole-shoes' (see page 332).

DISCRIMINATOR DISTORTION IN FREQUENCY-MODULATION SYSTEMS

By R. G. MEDHURST, B.Sc., and H. D. HYAMSON, B.E.E.

SUMMARY

In a previous paper[1] expressions were obtained for the distortion of the modulation when a frequency-modulated wave passes through a network whose group-delay and amplitude characteristics both exhibit small departures from constancy. An ideal discriminator network has the property that when it is fed with a frequency-modulated wave the amplitude modulation of the output is an exact copy (except for a scale factor) of the frequency modulation of the input. This can be achieved by a network whose group-delay characteristic is constant and whose amplitude characteristic is linear. In the present paper the theory given in the previous paper is extended to cover discriminator networks whose characteristics exhibit small departures from these ideal forms. Single-tone and noise-band modulations are considered. Two numerical examples are given, one being a design of 2-tuned-circuit discriminator, which is considered in relation to a 240-channel f.d.m. telephony system, and the other is an improved balanced discriminator suitable for a 600-channel system.

LIST OF PRINCIPAL SYMBOLS

ω_c = Carrier frequency, rad/sec.

ω_δ = Difference between a general angular frequency and ω_c, rad/sec.

ω_M = Frequency modulation, rad/sec.

ω_m = Frequency of one component tone of ω_M, rad/sec.

$\hat{\omega}_m$ = Maximum value of ω_m, rad/sec.

$\check{\omega}_m$ = Minimum value of ω_m, rad/sec.

ω_D = 'Peak' frequency deviation (taken as 11 dB above the r.m.s. deviation), rad/sec.

ω_a = Modulation frequency (single-tone modulation), rad/sec.

ω_d = Frequency deviation (single-tone modulation), rad/sec.

β = Modulation index = ω_d/ω_a (single-tone modulation).

ϕ_M = Phase modulation = $\int \omega_M dt$, rad.

$\phi_{\omega m}$ = Random phase angle associated with the tone of frequency ω_m, rad.

a = Amplitude of one component tone of ω_M, (rad/sec)$^{1/2}$, = $0\cdot40\omega_D\hat{\omega}_m^{-0\cdot5}$.

X = $\cos\phi_M$.

Y = $\sin\phi_M$.

V = Amplitude of frequency-modulated wave before transmission through discriminator, volts.

a_n = Coefficient of nth-order term in expansion of discriminator phase characteristic in powers of ω_δ/ω_c.

b_n = Coefficient of nth-order term in expansion of discriminator amplitude characteristic in powers of $\omega_\delta/\omega_c (n \geqslant 2)$.

B_0, B_1 = Coefficients respectively of zeroth and first-order terms in expansion of discriminator amplitude characteristic in powers of ω_δ/ω_c.

c_n = $B_0 a_n + B_1 a_{n-1}$.

A = $B_0 + \dfrac{B_1}{\omega_c}\omega_M$.

ρ = Non-linear portion of discriminator amplitude characteristic.

ϕ = Non-linear portion of discriminator phase characteristic.

$J_n(x)$ = Bessel function of first kind, of order n and argument x.

For the 2-tuned circuit discriminator, suffixes l and h denote circuits whose resonant frequencies are respectively lower and higher than the carrier frequency.

g_{ml} and g_{mh} = Respective mutual conductances of valves V_l and V_h (see Fig. 1).

v_g = Instantaneous amplitude of f.m. input wave, volts.

ω_{0l} and ω_{0h} = Resonant frequencies of the respective tuned circuits, rad/sec.

ω_0 = Resonant frequency of a single tuned circuit, rad/sec.

Q_l and Q_h = Q-factors of the respective tuned circuits.

C_l and C_h = Total capacitances across respective tuned circuits.

C = Total capacitance across a single tuned circuit.

η_l and η_h = Efficiencies of diodes D_l and D_h (see Fig. 1).

(1) INTRODUCTION

When a wave carrying frequency modulation but no amplitude modulation is passed through a network whose amplitude characteristic is accurately linear with frequency and whose group-delay characteristic is constant, the output is amplitude modulated by a precise copy (except for a scale factor) of the original frequency modulation.[2] This is the basis of certain well-known types of discriminator. In practice, both characteristics will depart from these ideal forms, and as a result non-linear distortion terms will appear in the outcoming amplitude modulation. The paper is concerned with the evaluation of these distortion terms.

In a previous paper,[1] an expression was obtained connecting the distortion of the phase modulation with that of the r.f. spectrum when a frequency-modulated carrier is passed through a passive network. With amplifiers, nearly matched aerial feeders, etc., in mind, the network was considered to be such that departures of its group-delay and amplitude characteristics from constancy (with respect to frequency) were both small. It will be shown that a straightforward extension can be made to the case of the discriminator network in which the group-delay characteristic exhibits small departures from constancy and the amplitude characteristic exhibits small departures from linearity. The expression to be obtained must now connect the distortion of the outcoming amplitude modulation (in so far as this is intended to be a perfect copy of the original frequency modulation) with the distortion of the r.f. spectrum.

As in the previous paper, it is found that, in the simple case of modulation by a single tone, harmonic distortion levels can be evaluated using numerical values of the characteristics at sideband positions, even though the characteristics may not be known analytically. For more complicated signals the theoretical situation is not so satisfactory. However, when the charac-

[357]

teristics can be described by power series, an expression for the distortion can be derived in terms of products of powers of the modulation and its time derivatives, and for characteristics in which the higher orders are not significant, this approach can be satisfactory. The case of a modulating signal consisting of a flat noise band (simulating an f.d.m. telephony signal[3]) is considered in some detail.

As an example, the design of the 2-tuned-circuit discriminator[8] is considered from the point of view of minimizing distortion. It is concluded that this simple type of discriminator may not be entirely satisfactory for use in f.d.m. frequency-modulation systems carrying large numbers of channels when performance to C.C.I.R. recommendations[4, 5] is envisaged. As a second example, distortion levels are derived for an improved balanced discriminator which is considered suitable for systems carrying up to 600 channels.

(2) RELATIONSHIP BETWEEN THE DISTORTION OF THE OUTPUT AMPLITUDE MODULATION AND THE SPECTRUM DISTORTION PRODUCED BY THE DISCRIMINATOR NETWORK

The undistorted wave entering the discriminator (supposed to be preceded by a perfect limiter) may be written as

$$V \sin (\omega_c t + \phi_M) = V \cos \phi_M \sin \omega_c t + V \sin \phi_M \cos \omega_c t \quad (1)$$

If it is known how to express $\cos \mu_t$ and $\sin \mu_t$ as sums of a number of tones, the right-hand side of eqn. (1) immediately yields the r.f. spectrum of the modulated wave. Consequently, we shall apply distorting characteristics to the right-hand side, and attempt to deduce the resulting distortion to the amplitude of the left-hand side.

Suppose that the network phase and amplitude characteristics are perfectly linear, being represented respectively by

$$a_0 + a_1 \frac{\omega_\delta}{\omega_c} \quad \text{and} \quad B_0 + \frac{B_1 \omega_\delta}{\omega_c}$$

The linear coefficient in the amplitude characteristic is written as a capital letter to emphasize that, in contrast with the theory given in Reference 1, this coefficient cannot be treated as small in a discriminator circuit. The term a_0 produces a constant phase-shift of all sidebands, and hence of the carrier only, without affecting the modulation of the emerging wave. The term $(a_1/\omega_c)\omega_\delta$ delays both the carrier and the modulation by an amount a_1/ω_c. Both these terms can be neglected if it is remembered that after passage through the network the quantity t appearing in the signal is actually $t_i + (a_1/\omega_c)$, where t_i is the time of entering. To deal with B_0 and B_1 we use a result due to Cherry and Rivlin,[2] who show that, when ϕ_M is any multitone signal, a network having an exactly linear amplitude characteristic and zero phase-shift acts as a perfect discriminator circuit, producing an output of the form

$$V \left[B_0 + \frac{B_1}{\omega_c} \omega_M \right] \sin (\omega_c t + \phi_M)$$
$$= VA \sin (\omega_c t + \phi_M)$$
$$= VA \cos \phi_M \sin \omega_c t + VA \sin \phi_M \cos \omega_c t \quad (2)$$

Now, suppose that the phase and amplitude characteristics each depart from linearity by a small amount. The expression for the output in eqn. (2) will become

$$V(A \cos \phi_M + \Delta_1) \sin \omega_c t + V(A \sin \phi_M + \Delta_2) \cos \omega_c t$$

where Δ_1 and Δ_2 are distortion terms. If the corresponding distortion of the amplitude modulation is Δ,

$$V(A + \Delta) \sin (\omega_c t + \phi_M + \theta) = V(A \cos \phi_M + \Delta_1) \sin \omega_c t$$
$$+ V(A \sin \phi_M + \Delta_2) \cos \omega_c t$$

where θ is a phase-distortion term. Thus,

$$(A + \Delta)^2 = (A \cos \phi_M + \Delta_1)^2 + (A \sin \phi_M + \Delta_2)^2$$

so that, to first order,

$$\Delta = \Delta_1 \cos \phi_M + \Delta_2 \sin \phi_M \quad \cdot \quad \cdot \quad \cdot$$

It will be noticed that this is very similar in form to the expression derived in Reference 1, which is, in the present notation,

$$\Delta = \Delta_2 \cos \phi_M - \Delta_1 \sin \phi_M \quad \cdot \quad \cdot \quad \cdot \quad (3a)$$

An important difference, which has to be remembered when considering the combined effect of distortion due to amplifier circuits, etc., and to discriminator circuits, is that eqn. represents distortion of the *frequency* modulation (after it has been transferred to the amplitude of the modulated wave), while eqn. (3a) represents distortion of the *phase* modulation.

(3) MODULATION BY A SINGLE TONE

The modulated wave may be written as

$$\sin (\omega_c t + \beta \sin \omega_a t)$$

where the amplitude has been made unity.

The amplitude and phase characteristics of the discriminator network will be taken to be respectively $B_0 + (B_1/\omega_c)\omega_\delta + \rho$ and ϕ, where ρ and ϕ are each functions of frequency. Both characteristics are assumed to depart from linearity only to a small extent over the effective band. Thus, ρ may be considered to be small compared with $B_0 + (B_1/\omega_c)\omega_\delta$. It has already been remarked in Section 2 that constant and linear phase terms produce no distortion: the former merely alters the phase of the carrier and the latter can be accounted for by introducing a suitable delay into the expressions for carrier and modulation. Consequently, it can be assumed that sufficient constant and linear terms have been subtracted from the phase characteristic to ensure that ϕ does not exceed a small fraction of a radian over the r.f. band. Thus, second-order terms in ρ and ϕ can be neglected.

The detailed analysis is similar to that given in Section 3 of Reference 1, and will consequently not be repeated here. In the result, it is found that harmonic distortion ratios can be calculated with the help of a Table of distortion factors, which can be obtained by making the following modifications to Table of Reference 1. The entries in this Table, in the columns headed 'Phase characteristic', have to be multiplied by $\beta/2$ for second-harmonic distortion and by $\beta/3$ for third-harmonic distortion in order to obtain entries in the new Table which are to be used in conjunction with the discriminator amplitude characteristic. The entries to be used in conjunction with the (modified) discriminator phase characteristic are similarly obtained from entries in the original Table in the columns headed 'Amplitude characteristic', with a reversal of signs. Second- and third-harmonic distortion of single-tone modulation can now be found from this Table by multiplying each of the values of ρ and $[B_0 + (B_1/\omega_c)n\omega_a]\phi$ at each of the sideband positions by the appropriate entry, summing the resulting products, and dividing by $B_1\omega_d/\omega_c$.

(4) MODULATION BY A GENERAL SIGNAL: CHARACTERISTICS EXPRESSED AS POWER SERIES

When the phase and amplitude characteristics are expressed as power series with respect to frequency, it becomes possible to evaluate the distortion of a general signal within certain, quite wide, restrictions. This is of practical importance when the characteristics can be represented sufficiently closely by the first few terms of their series expansions. Such is the case, for

ample, when the network is a resonant circuit and the carrier frequency is situated on the more-or-less linear part of the amplitude characteristic, the frequency excursion being not too large.

The conditions that need to be satisfied are that all differential coefficients of ϕ_M with respect to time exist, and that $\cos\phi_M$ and $\sin\phi_M$ can be represented by some suitable number of tones. Let the amplitude and phase characteristics be respectively

$$\ldots + \frac{B_1}{\omega_c}\omega_\delta + b_0 + \frac{b_1}{\omega_c}\omega_\delta + \frac{b_2}{\omega_c^2}\omega_\delta^2 + \frac{b_3}{\omega_c^3}\omega_\delta^3 +$$

$$\ldots = B_0 + \frac{B_1\omega_\delta}{\omega_c} + \rho(\omega_c + \omega_\delta)$$

and

$$a_0 + \frac{a_1}{\omega_c}\omega_\delta + \frac{a_2}{\omega_c^2}\omega_\delta^2 + \frac{a_3}{\omega_c^3}\omega_\delta^3 \ldots = \phi(\omega_c + \omega_\delta)$$

b_0 and b_1 being supposed to be very much less than B_0 and B_1 respectively. They are inserted for analytical completeness, although in practice (e.g. in the case of a characteristic exactly known, such as that in Section 6) they could be absorbed into b_0 and B_1. It is assumed, as before, that the major portions of the constant and linear terms have been subtracted from ϕ.

Then it can be shown (Section 8.1) that the distortion appearing in the amplitude modulation of the outcoming wave is

$$b_0(X^2 + Y^2) + \frac{b_1}{\omega_c}(XY^I - X^I Y) - \frac{b_2}{\omega_c^2}(XX^{II} + YY^{II})$$

$$- \frac{b_3}{\omega_c^3}(XY^{III} - YX^{III}) + \frac{b_4}{\omega_c^4}(XX^{IV} + YY^{IV})$$

$$+ \frac{b_5}{\omega_c^5}(XY^V - YX^V) - \ldots + \frac{c_1}{\omega_c}(XX^I + YY^I)$$

$$+ \frac{c_2}{\omega_c^2}(XY^{II} - YX^{II}) - \frac{c_3}{\omega_c^3}(XX^{III} + YY^{III})$$

$$- \frac{c_4}{\omega_c^4}(XY^{IV} - YX^{IV}) + \frac{c_5}{\omega_c^5}(XX^V + YY^V) + \ldots \quad (4)$$

where the superscripts denote differentiation with respect to time. Explicitly, the first few terms of the amplitude modulation are as follows:

$$\ldots + b_0 + \frac{(B_1 + b_1)}{\omega_c}\omega_M + \frac{b_2}{\omega_c^2}(\omega_M)^2 + \frac{b_3}{\omega_c^3}(\omega_M)^3$$

$$+ \frac{b_4}{\omega_c^4}[(\omega_M)^4 - 3(\omega_M^I)^2 - 4\omega_M\omega_M^{II}]$$

$$+ \frac{b_5}{\omega_c^5}[(\omega_M)^5 - 15\omega_M(\omega_M^I)^2 - 10(\omega_M)^2\omega_M^{II}] + \ldots$$

$$+ \frac{c_2}{\omega_c^2}\omega_M^I + 3\frac{c_3}{\omega_c^3}\omega_M\omega_M^I + \frac{c_4}{\omega_c^4}[6(\omega_M)^2\omega_M^I - \omega_M^{III}]$$

$$+ \frac{c_5}{\omega_c^5}[10(\omega_M)^3\omega_M^I - 10\omega_M^I\omega_M^{II} - 5\omega_M\omega_M^{III}] + \ldots \quad (5)$$

ω_M being here taken as a function of $(t + A_1/\omega_c)$, where A_1/ω_c is the coefficient of the linear term subtracted from the phase characteristic.

Distortion products associated with the phase characteristic are seen to be in quadrature with products of the same frequency associated with the amplitude characteristic.

It does not appear possible to derive the leading terms of expression (5) from quasi-stationary arguments alone. The reason for this can be seen by treating the discriminator circuit as two successive stages, the first having the phase characteristic

of the original circuit and unity amplitude characteristic, and the second having the amplitude characteristic of the original circuit and zero phase-shift. Then the leading terms of the expression for the distortion of the final outcoming amplitude modulation depend in part on the amplitude modulation generated in the wave emerging from the first stage, which is not predictable by quasi-stationary approach.

If the characteristics of the amplifier circuits preceding the discriminator can also be expressed as power series, the results of the present section can be combined with those of Reference 1 to give the total distortion due to a system consisting of an amplifier and a discriminator with intervening perfect limiter. It is then seen that the distortions due to the phase characteristics of the amplifier and discriminator are in phase, as are those due to the amplitude characteristics. Thus some compensation of distortion due to a characteristic of one of these system elements may be achieved by suitable choice of the appropriate characteristic of the other. However, distortion terms of the same order turn out in the two cases to be quite different functions of ω_M and its derivatives, so that complete cancellation does not seem possible when ω_M is made up of more than one tone.

(5) NOISE-BAND MODULATION

For test purposes it is customary to simulate an f.d.m. telephony signal by a band of random noise having a flat power spectrum extending between the frequencies occupied by the upper and lower speech channels. When this is applied as a frequency modulation we can write ω_M in the form

$$\omega_M = a \sum_{\omega_m = \check{\omega}_m}^{\hat{\omega}_m} \cos(\omega_m t + \phi\omega_m) \quad . \quad . \quad (6)$$

where ω_m is supposed to increase in steps of 1 rad/sec.

$\check{\omega}_m$ will in general be much less than $\hat{\omega}_m$, and it will be sufficiently accurate for the present purpose to take it as zero (or unity, whichever is convenient).

A derivation of the non-linear distortion generated by the terms up to b_5 and c_4 in expression (5) is given in outline in Section 8.2. General expressions applying at any base-band frequency are quite lengthy. It will be sufficient here to give formulae relating to the upper end of the base-band, where distortion tends to be worst, and these are given in Table 1.

(6) EXAMPLES

(6.1) The 2-Tuned-Circuit Discriminator

The 2-tuned-circuit discriminator, shown in Fig. 1, is widely used in trunk radio systems. In this arrangement, valves V_1 (mutual conductance g_{ml}) and V_h (mutual conductance g_{mh}) feed, from a common input signal v_g (angular carrier frequency ω_c), two parallel LCR circuits, of resonant angular frequencies $\omega_{0l}(\omega_{0l}/\omega_c < 1)$ and $\omega_{0h}(\omega_{0h}/\omega_c > 1)$, and Q-factors Q_l and Q_h. The total capacitances across each circuit are C_l and C_h, and diodes D_l and D_h, of efficiencies η_l and η_h respectively, detect the amplitude modulation produced by the parallel LCR circuits when v_g is an f.m. wave. With the diodes connected as shown in Fig. 1, stage S subtracts the diode outputs to give the resultant discriminator output.

Various sources of distortion are:

  (a) Non-linearity of the amplitude characteristics of the tuned circuits.
  (b) Non-linearity of the phase characteristics of the tuned circuits.
  (c) Amplitude modulation of the input wave.
  (d) Harmonics of the carrier appearing in the anode circuits of the valves V_1 and V_h due to non-linearity of the valve characteristics or to harmonics in the input wave.
  (e) Non-linearity of the characteristics of the diodes D_l and D_h.

Table 1

EXPRESSIONS FOR DISTORTION/SIGNAL IN TOP CHANNEL (NOISE-BAND FREQUENCY MODULATION)

Distortion mechanism	Distortion/signal in top channel (voltage ratio)
Second-order, due to amplitude characteristic	$0 \cdot 28 \dfrac{\omega_D}{\omega_c} \dfrac{b_2}{B_1} \left\{ 1 + \left(\dfrac{\omega_D}{\omega_c}\right)^2 \dfrac{b_4}{b_2} \left[0 \cdot 96 + \dfrac{11}{3}\left(\dfrac{\hat{\omega}_m}{\omega_D}\right)^2 \right] \right.$ $\left. + \left(\dfrac{\omega_D}{\omega_c}\right)^4 \left(\dfrac{b_4}{b_2}\right)^2 \left[0 \cdot 23 + 1 \cdot 76\left(\dfrac{\hat{\omega}_m}{\omega_D}\right)^2 + \dfrac{101}{30}\left(\dfrac{\hat{\omega}_m}{\omega_D}\right)^4 \right] \right\}^{1/2}$
Third-order, due to amplitude characteristic	$0 \cdot 04 \left(\dfrac{\omega_D}{\omega_c}\right)^2 \dfrac{b_3}{B_1} \left\{ 12 + \left(\dfrac{\omega_D}{\omega_c}\right)^2 \dfrac{b_5}{b_3} \left[19 \cdot 2 + 5\left(\dfrac{\hat{\omega}_m}{\omega_D}\right)^2 \right] \right.$ $\left. + \left(\dfrac{\omega_D}{\omega_c}\right)^4 \left(\dfrac{b_5}{b_3}\right)^2 \left[7 \cdot 68 + 4\left(\dfrac{\hat{\omega}_m}{\omega_D}\right)^2 + \dfrac{470}{9}\left(\dfrac{\hat{\omega}_m}{\omega_D}\right)^4 \right] \right\}^{1/2}$
Fourth-order, due to amplitude characteristic	$0 \cdot 077 \left(\dfrac{\omega_D}{\omega_c}\right)^3 \dfrac{b_4}{B_1}$
Fifth-order, due to amplitude characteristic	$0 \cdot 47 \left(\dfrac{\omega_D}{\omega_c}\right)^4 \dfrac{b_5}{B_1}$
Second-order, due to phase characteristic	$0 \cdot 42 \left(\dfrac{\omega_D}{\omega_c}\right)^2 \left(\dfrac{\hat{\omega}_m}{\omega_D}\right) \dfrac{c_3}{B_1}$
Third-order, due to phase characteristic	$0 \cdot 28 \left(\dfrac{\omega_D}{\omega_c}\right)^3 \left(\dfrac{\hat{\omega}_m}{\omega_D}\right) \dfrac{c_4}{B_1}$

We shall consider only quantitatively distortion due to (a) and (b), although the choice of tuned-circuit parameters is to some extent influenced by the desirability of minimizing distortion from the other sources mentioned. There are five independent parameters available, namely ω_{0l} and ω_{0h}, Q_l and Q_h, and C_l/C_h. This latter ratio determines the relative voltages applied to the two diodes, which may also be controlled by adjustment of the ratio g_{ml}/g_{mh}.

A detailed discussion of the considerations leading to a choice of parameter values would be out of place in the present paper. The following much abbreviated outline is offered to indicate the line of thought adopted.

(6.1.1) Basic Formulae.

The modulus of the impedance Z of one tuned circuit, at angular frequency $\omega_c + \omega_\delta$, is given by

$$|Z|\omega_c C = \frac{Q\omega_c/\omega_0}{\sqrt{\left[1 + Q^2\left(\dfrac{\omega_0}{\omega_c + \omega_\delta} - \dfrac{\omega_c + \omega_\delta}{\omega_0}\right)^2\right]}} \quad . \quad (7)$$

where $Q = \omega_0 CR$, and the suffixes l and h have been omitted. The choice of $1/\omega_c C$ as a normalizing factor somewhat simplifies the detailed analysis.

The phase angle ϕ is given by

$$\phi = \arctan\left[Q\left(\dfrac{\omega_0}{\omega_c + \omega_\delta} - \dfrac{\omega_c + \omega_\delta}{\omega_0}\right) \right] \quad . \quad . \quad (8)$$

The right-hand sides of eqns. (7) and (8) have to be expanded in powers of ω_δ, in order to obtain the coefficients required in eqn. (5). This is straightforward in principle, although the formulae increase in complexity as the orders of the coefficients increase. These coefficients are functions of Q and ω_c/ω_0, and the final discriminator design is conveniently based on families of curves showing values of each coefficient plotted against ω_c/ω_0, for various Q-factors.

Then, from eqn. (5), the envelope of the voltage across the
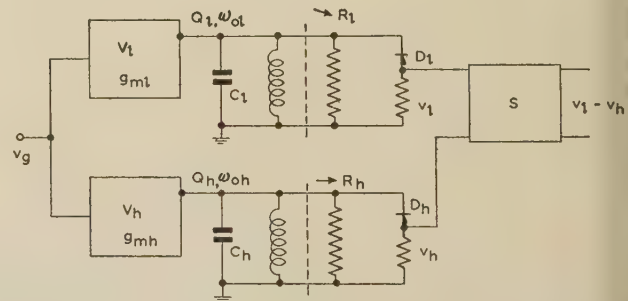
Fig. 1.—Schematic representation of 2-tuned-circuit discriminator.

tuned circuit when fed from an f.m. constant-current source $I = |I| \cos(\omega_c t + \phi_M)$ is

$$|v| = \frac{|I|}{\omega_c C}\left\{ B_0 + \frac{B_1}{\omega_c}\omega_M + \frac{b_2}{\omega_c^2}(\omega_M)^2 + \frac{b_3}{\omega_c^3}(\omega_M)^3 \right.$$

$$+ \frac{b_4}{\omega_c^4}[(\omega_M)^4 - 3(\omega_M^I)^2 - 4\omega_M\omega_M^{II}] + \ldots$$

$$+ \frac{c_2}{\omega_c^2}\omega_M^I + 3\frac{c_3}{\omega_c^3}\omega_M\omega_M^I + \frac{c_4}{\omega_c^4}[6(\omega_M)^2\omega_M^I - \omega_M^{III}]$$

$$\left. + \frac{c_5}{\omega_c^5}[10(\omega_M)^3\omega_M^I - 10\omega_M^I\omega_M^{II} - 5\omega_M\omega_M^{III}] + \ldots \right.$$

$$. \quad . \quad . \quad (9)$$

The term in c_2 is not of importance in the transmission of multiplex telephony, since it contributes no intermodulation distortion. It does, however, give rise to a varying phase-shift across the video band, and this would cause distortion of television transmission.

The mean amplitude of the voltage across the tuned circuit is given by the B_0 term in eqn. (9), the modulation signal component

given by the B_1 term, and, in general, distortion components of the modulation signal in the envelope arise from all the other terms.

If the constant-current source is a pentode of mutual conductance g_m,

$$|I| = g_m|v_g|$$

where $|v_g|$ is the amplitude of the grid-cathode p.d., henceforth referred to as the discriminator input. Assuming that the detection is linear, the output from the tuned 2-circuit discriminator is given by

$$\frac{\eta|v_g|}{\omega_c}\left\{\left(\frac{B_{0l}}{C_l} - \frac{B_{0h}}{C_h}\right) + \left(\frac{B_{1l}}{C_l} - \frac{B_{1h}}{C_h}\right)\frac{\omega_M}{\omega_c} + \left(\frac{b_{2l}}{C_l} - \frac{b_{2h}}{C_h}\right)\frac{(\omega_M)^2}{\omega_c^2}\right.$$
$$+ \left(\frac{b_{3l}}{C_l} - \frac{b_{3h}}{C_h}\right)\frac{(\omega_M)^3}{\omega_c^3} + \left(\frac{b_{4l}}{C_l} - \frac{b_{4h}}{C_h}\right)\left[\frac{(\omega_M)^4}{\omega_c^4} + \ldots\right]$$
$$\left. + \ldots + 3\left(\frac{c_{3l}}{C_l} - \frac{c_{3h}}{C_h}\right)\frac{\omega_M\omega_M^l}{\omega_c^3} + \ldots\right\} \quad . \quad . \quad . \quad (10)$$

in which it has been assumed that

$$g_m = g_{ml} = g_{mh}$$

and

$$\eta = \eta_l = \eta_h$$

B_{1l} and B_{1h} are of opposite sign if $\omega_{0l}/\omega_c < 1 < \omega_{0h}/\omega_c$. The difference in group delay, $(a_{1l} - a_{1h})/\omega_c$, of the tuned circuits at the carrier frequency has been neglected in writing eqn. (10). When this difference is taken into account, the coefficients in the equation no longer have the form of simple subtractions but must be obtained by compounding two vectors not quite in anti-phase. A discriminator design which relies on cancellation between the two sides must, of course, make suitable allowance for this effect. For the design given below,

$$\frac{a_{1l}}{\omega_c} - \frac{a_{1h}}{\omega_c} = 2 \cdot 14 \times 10^{-9}\,\text{sec}$$

so that at the top of the base-band ($1 \cdot 052$ Mc/s in a 240-channel system), the phase difference is about $0 \cdot 8°$. This particular design is not based on cancellation, so that the modification to the coefficients in eqn. (10) can be considered negligible.

(6.1.2) **Determination of Parameters.**

Two initial requirements, which use two parameters, are

$$\frac{B_{0l}}{C_l} = \frac{B_{0h}}{C_h} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and

$$\frac{B_{1l}}{C_l} = -\frac{B_{1h}}{C_h} \quad . \quad . \quad . \quad . \quad (12)$$

Here, as in the previous Section, the B's are coefficients in the expansion of the right-hand side of eqn. (7). The purpose of eqn. (11) is to minimize distortion due to amplitude modulation of the input signal (it being assumed that $\eta_l = \eta_h$ and $g_{ml} = g_{mh}$). Eqns. (11) and (12) together ensure that the operating conditions of the two diodes are similar, with the object of reducing, through cancellation in the stage S (Fig. 1), even-order distortion arising from the two diodes.

The remaining three parameters of the two tuned circuits are now chosen with the aim of minimizing distortion arising from the circuits themselves. Two alternative courses are

(a) To make each of the contributions of the two tuned circuits to one or more of the coefficients in eqn. (10) separately zero.
(b) To make the contributions equal and of the same sign, in an attempt to achieve cancellation.

The performance of designs based on each of these approaches depends on the stability of circuit elements and of frequency, but the performance of a design based on course (b) also involves the stability of g_{mh}, g_{ml}, η_h and η_l; course (a) has therefore been taken. Only one source of distortion can be dealt with in this way, since two parameters are required to minimize each distortion source. The best choice appears to be the second-order distortion associated with the amplitude characteristic, since this tends to be the highest component generated by each tuned circuit. Thus, the resonant frequencies are chosen so that the carrier frequency lies at the point of inflection of each circuit.

There remains one parameter, in terms of which all the others can be expressed. Let $\omega_{0h}/\omega_c = y$, say, be this parameter, and put $\omega_{0l}/\omega_c = x$. The requirement that the second-order amplitude characteristic shall be zero imposes the condition

$$b_{2l} = b_{2h} = 0 \quad . \quad . \quad . \quad . \quad . \quad (13)$$

and it can be shown that, from eqns. (11)–(13),

$$x^4 = \frac{y^4 + 1}{3y^4 - 1} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

$$\left(\frac{C_l}{C_h}\right)^2 = \frac{3y^4 - 1}{2} \quad . \quad . \quad . \quad . \quad (15)$$

$$Q_h^2 = \frac{y^2(1 + 3y^4)}{2(1 - y^2)^2(1 + 3y^2)} \quad . \quad . \quad . \quad (16)$$

$$Q^2 = \frac{x^2(1 + 3x^4)}{2(1 - x^2)^2(1 + 3x^2)} \quad . \quad . \quad . \quad (17)$$

$$B_{0h} = \frac{(1 + 3y^4)^{1/2}}{(y^4 - 1)\sqrt{3}} \quad . \quad . \quad . \quad (18)$$

$$B_{1h} = \frac{(1 + 3y^4)^{3/2}}{(y^4 - 1)^2 3\sqrt{3}} \quad . \quad . \quad . \quad (19)$$

The choice of y depends on a number of conflicting considerations. The third-order amplitude characteristic distortion (and hence any second-order amplitude characteristic distortion due to a small shift of the carrier frequency), and also the fourth-order amplitude-characteristic distortion and the second-order phase-characteristic distortion all decrease as y increases, although the rate of decrease is slow when y exceeds $1 \cdot 3$. Distortion due both to amplitude modulation of the incoming wave and to non-linearity of diode characteristics is expected to increase as y increases. Quantitative consideration of these effects suggests $y = 1 \cdot 25$ as a good compromise value.

(6.1.3) **Distortion Levels.**

The coefficients required for evaluating distortion components from the formulae given in Table 1 can now be determined. They are as follows:

$$b_2/B_1 = -0 \cdot 0358$$
$$b_3/B_1 = -7 \cdot 15$$
$$b_4/B_1 = -14 \cdot 75$$
$$c_3/B_1 = -3 \cdot 67$$

The non-zero value of b_2/B_1 follows from the assumption of a drift of $0 \cdot 1$ Mc/s in the carrier frequency, which is taken in this example as 60 Mc/s.

Suppose that the maximum modulating frequency, $\hat{\omega}_m/2\pi$, is $1 \cdot 052$ Mc/s and the peak deviation $\omega_D/2\pi$, is $2 \cdot 3$ Mc/s (corresponding to a 240-channel system in which the r.m.s. deviation per channel is 163 kc/s). Then the distortion levels calculated from Table 1 are given in Table 2.

Table 2

DISTORTION LEVELS FOR A 2-TUNED-CIRCUIT DISCRIMINATOR DESIGNED AS IN SECTION 7 (240-CHANNEL SYSTEM)

Distortion mechanism	Distortion/signal ratio in top channel	Absolute distortion level in top 4 kc/s channel
	dB	dBmo
Second-order, due to amplitude characteristic	−64·7	−76·5
Third-order, due to amplitude characteristic	−56·6	−68·4
Fourth-order, due to amplitude characteristic	−83·9	−95·7
Second-order, due to phase characteristic	−59·7	−71·5
Total	−54·4	−66·2

The trend of the distortion levels suggests that higher-order terms in both characteristics will contribute negligibly little distortion, but this cannot at present be verified directly for noise-band modulation. Using single-tone modulation it is possible to work out harmonic levels due to the first few terms of each characteristic, and also, from Section (3), those due to the complete characteristic. When the tone is inserted at the same r.m.s. level as the noise band, it can be verified that, for single-tone modulating frequencies lying within the modulating noise-band frequency range, both second- and third-harmonic distortion are dominated by the mechanisms listed in Table 2.

The distortion allocation for a single demodulator depends on C.C.I.R. requirements[4, 5] for a complete system, and also on the distribution of this total distortion among the various units of the system. This is a complicated matter which cannot be gone into here. For purpose of comparison with the level shown at the end of Table 2, it will merely be stated that for a 2 500 km system a distortion breakdown thought to accord as well as possible with existing data leads to an allocation of −69 dBmo distortion in the top 4 kc/s channel per demodulator.

Thus it appears that, at least in the form developed here, the 2-tuned-circuit discriminator will be somewhat inadequate for 240-channel systems. For systems carrying larger numbers of channels the performance of this type of discriminator will fall even further short of the requirement, and more complicated networks must be considered. As a second example, the following Section treats one form of balanced discriminator proposed for 600-channel systems.

(6.2) Improved Discriminator Design

Circuit diagrams of the two halves of this balanced discriminator[7] are shown in Fig. 2.

The amplitude and phase characteristics, expressed as power series in δ (i.e. ω_s/ω_c), are as follows:

Positive-slope network.

Amplitude characteristic (normalized with respect to the coefficient of δ):

$$0\cdot214 + \delta + 0\cdot000\delta^2 + 0\cdot000\delta^3 − 0\cdot000\delta^4$$
$$− 0\cdot000\delta^5 − 473\cdot7\delta^6 \cdots$$

Phase characteristic (radians):
$$− 1\cdot336 − 8\cdot198\delta + 0\cdot181\delta^2 − 23\cdot081\delta^3 + 2\cdot066\delta^4 \cdots$$

Negative-slope network.

Amplitude characteristic (normalized with respect to the coefficient of δ):

$$0\cdot214 − \delta + 0\cdot000\delta^2 − 0\cdot000\delta^3 + 0\cdot000\delta^4$$
$$− 0\cdot000\delta^5 − 475\cdot4\delta^6 \cdots$$

Phase characteristic (radians):
$$− 1\cdot319 − 8\cdot193\delta − 0\cdot109\delta^2 − 22\cdot814\delta^3 − 0\cdot450\delta^4 \cdots$$
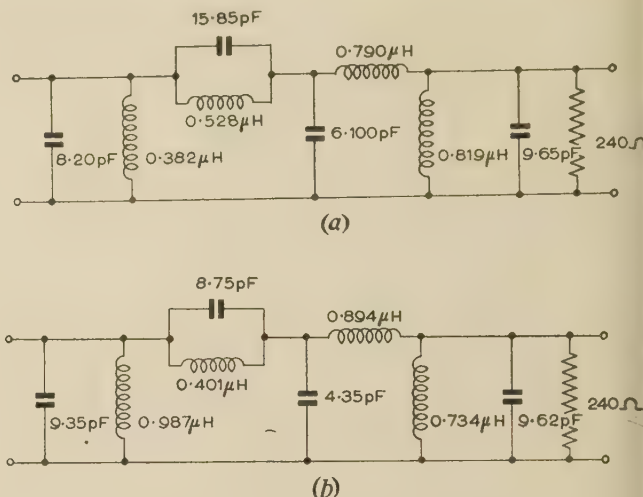
Fig. 2.—Improved discriminator design for 70 Mc/s i.f. stage.
(a) Positive-slope network.
(b) Negative-slope network.

It will be noticed that, to the accuracy shown, the coefficients of the second-, third-, fourth- and fifth-order terms in the amplitude characteristics disappear. This is, in fact, a design basis of these networks.

Distortion levels derived from the formulae in Section 5 are as given in Table 3.

Table 3

DISTORTION LEVELS FOR POSITIVE- AND NEGATIVE-SLOPE NETWORKS

Distortion mechanism	Distortion level in top 4 kc/s channel	
	Positive-slope network	Negative-slope network
	dBmo	dBmo
Second-order, due to phase characteristic	−60·5	−60·7
Third-order, due to phase characteristic	−75·4	−75·4
Second-order, due to amplitude characteristic (sixth-power term)	−90·0	−90·0

The signs of the coefficients are such that subtraction of the outputs from the two sides of the discriminator results in cancellation of the main term of the second-order phase-characteristic distortion and reinforcement of the main term of the third-order phase-characteristic distortion. If it is assumed that, in practice, a 14 dB reduction of the main second-order distortion term (and hence a 20 dB reduction in distortion/signal ratio, since the signal voltage is itself doubled) can be achieved in this way, we arrive at a residual second-order distortion of −80·5 dBmo. Thus, the total distortion in the discriminator output due to circuit non-linearity comes out as −74·3 dBmo—some 5 dB better than required for multi-channel working.

(7) REFERENCES

(1) MEDHURST, R. G.: 'Harmonic Distortion of Frequency Modulated Waves by Linear Networks', *Proceedings I.E.E.*, Paper No. 1650 R, May, 1954 (**101**, Part II, p. 171).

(2) CHERRY, E. C., and RIVLIN, R. S.: 'Non-Linear Distortion with Particular Reference to the Theory of Frequency Modulated Waves—Part II', *Philosophical Magazine*, 1942, **33**, p. 272.

(3) ALBERSHEIM, W. J., and SCHAFER, J. P.: 'Echo Distortion in the FM Transmission of Frequency-Division-Multiplex', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 316.

(4) C.C.I.R. Documents of the VIIth Plenary Assembly, London, 1953, Vol. 1 (Recommendation No. 40, p. 35).

(5) Tome III bis. C.C.I.F. XVIe Assemblée Plénière. Firenze, October, 1951.

(6) LEWIN, L.: 'Interference in Multi-Channel Circuits: Dependence on Harmonic Generation', *Wireless Engineer*, 1950, **27**, p. 294.

(7) EASTER, B.: 'An Approach to the Design of Frequency Discriminator Networks of High Linearity' (in preparation).

(8) FANCOURT, K. G., and SKWIRZYNSKI, J. K.: 'Design of a Simple Linear Frequency Discriminator', *Marconi Review*, 1956, **19**, p. 61.

(8) APPENDICES

(8.1) Form of the Distortion when the Network Characteristics can be represented by Power Series

Eqn. (4) can be obtained directly from Cherry and Rivlin's[2] expression for the r.f. spectrum of a frequency-modulated wave. In their notation (except that ω_c is used for the carrier frequency, in place of their ω_0),

$$V \sin \left[\omega_c t + \sum_{s=1}^{S} m_s \cos (\omega_s t + \epsilon_s) \right]$$

$$= \sum_{(\alpha_s)} V \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right] \cos \left[\omega_c t + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \phi_s \right] \quad (20)$$

where $\phi_s = \omega_s t + \epsilon_s$, $\alpha = \sum_{s=1}^{S} \alpha_s$, and the summation Σ has to be carried out over all possible sets of values of (α_s) where α_s may have any positive or negative integral value, or be zero.

Writing δ_s for $\sum_{s=1}^{S} \alpha_s \omega_s$, the right-hand side of (20) becomes

$$\sum_{(\alpha_s)} V \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right] \cos \left[(\omega_c + \delta_s)t + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \epsilon_s \right]$$

We assume, as in Section 3, that the amplitude and phase characteristics are respectively $B_0 + (B_1/\omega_c)\omega_s + \rho$ and ϕ, the symbols having the same meaning as in that Section. Then, the spectrum of the wave emerging from the network is

$$\sum_{(\alpha_s)} V \left[\sum_{s=1}^{S} J_{\alpha_s}(m_s) \right] \left[B_0 + \frac{B_1}{\omega_c} \delta_s + \rho(\omega_c + \delta_s) \right] \cos \left[(\omega_c + \delta_s)t \right.$$
$$\left. + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \epsilon_s + \phi(\omega_c + \delta_s) \right]$$

Expanding to first order in ρ and ϕ, this becomes

$$\sum_{(\alpha_s)} V \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right] \left[B_0 + \frac{B_1}{\omega_c} \delta_s \right] \cos \left[(\omega_c + \delta_s)t \right.$$
$$\left. + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \epsilon_s \right] + \sum_{(\alpha_s)} V \left[\prod_{s=1}^{S} J_{a_s}(m_s) \right]$$
$$\rho(\omega_c + \epsilon_s) \cos \left[(\omega_c + \delta_s)t + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \epsilon_s \right]$$

$$- \sum_{(\alpha_s)} V \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right] \left[B_0 + \frac{B_1}{\omega_c} \delta_s \right]$$

$$\phi(\omega_c + \delta_s) \sin \left[(\omega_c + \delta_s)t + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \omega_s \right] \quad (21)$$

The first summation represents the spectrum of a frequency-modulated wave which has passed through a network having linear amplitude characteristic and zero phase-shift. Cherry and Rivlin show that this can be written as

$$V \left[B_0 + \frac{B_1}{\omega_c} \phi'_M \right] \sin \left[\omega_c t + \phi_M \right]$$

where ϕ_M is the phase modulation, i.e. $\sum_{s=1}^{S} m_s \cos (\omega_s t + \epsilon_s)$, and the dash denotes differentiation with respect to time.

Suppose that ρ and ϕ can be expressed as power series of the form

$$b_0 + \frac{b_1}{\omega_c} \delta_s + \frac{b_2}{\omega_c^2} \delta_s^2 + \dots$$

and

$$a_0 + \frac{a_1}{\omega_c} \delta_s + \frac{a_2}{\omega_c^2} \delta_s^2 + \dots$$

respectively, and put

$$\left(B_0 + \frac{B_1}{\omega_c} \delta_s \right) \left(a_0 + \frac{a_1}{\omega_c} \delta_s + \frac{a_2}{\omega_c^2} \delta_s^2 + \dots \right)$$
$$= c_0 + \frac{c_1}{\omega_c} \delta_s + \frac{c_2}{\omega_c^2} \delta_s^2 + \dots$$

Then, expression (21) may be written in the form

$$V \left[B_0 + \frac{B_1}{\omega_c} \phi'_M \right] \sin (\omega_c t + \phi_M) + \sin \omega_c t \left\{ -\sum_{(\alpha_s)} V \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right] \right.$$

$$\left[b_0 + \frac{b_1}{\omega_c} \delta_s + \frac{b_2}{\omega_c^2} \delta_s^2 \dots \right] \sin \left[\delta_s t + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \epsilon_s \right]$$

$$- \sum_{(\alpha_s)} V \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right] \left[c_0 + \frac{c_1}{\omega_c} \delta_s + \frac{c_2}{\omega_c^2} \delta_s^2 + \dots \right] \cos \left[\delta_s t + (\alpha - 1)\frac{\pi}{2} \right.$$

$$\left. + \sum_{s=1}^{S} \alpha_s \epsilon_s \right] \right\} + \cos \omega_c t \left\{ \sum_{(\alpha_s)} V \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right] \left[b_0 + \frac{b_1}{\omega_c} \delta_s \right. \right.$$

$$\left. + \frac{b_2}{\omega_c^2} \delta_s^2 \dots \right] \cos \left[\epsilon_s t + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \epsilon_s \right] - \sum_{(\alpha_s)} V \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right]$$

$$\left. \left[c_0 + \frac{c_1}{\omega_c} \delta_s + \frac{c_2}{\omega_c^2} \delta_s^2 + \dots \right] \sin \left[\delta_s t + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \epsilon_s \right] \right\}$$
$$\qquad \qquad \qquad \qquad \cdots \quad (22)$$

Put
$$X = \cos \phi_M = - \sum_{(\alpha_s)} \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right] \sin \left[\delta_s t + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \epsilon_s \right]$$

and
$$Y = \sin \phi_M = \sum_{(\alpha_s)} \left[\prod_{s=1}^{S} J_{\alpha_s}(m_s) \right] \cos \left[\delta_s t + (\alpha - 1)\frac{\pi}{2} + \sum_{s=1}^{S} \alpha_s \epsilon_s \right]$$

(Reference 1, Section 9.4).

Then, it is readily seen that expression (22) may be rewritten as

$$V[B_0 + B_1 \phi'_M] \sin (\omega_c t + \phi_M)$$

$$+ V \sin \omega_c t \left[b_0 X + \frac{b_1 Y^I}{\omega_c} - \frac{b_2 X^{II}}{\omega_c^2} - \frac{b_3 Y^{III}}{\omega_c^3} + \frac{b_4 X^{IV}}{\omega_c^4} + \dots \right.$$

$$- c_0Y + \frac{c_1X^I}{\omega_c} + \frac{c_2X^{II}}{\omega_c^2} - \frac{c_3X^{III}}{\omega_c^3} - \frac{c_4X^{IV}}{\omega_c^4} + \cdots \Bigg]$$

$$+ V\cos\omega_c t \Bigg[b_0Y - \frac{b_1X^I}{\omega_c} - \frac{b_2Y^{II}}{\omega_c^2} + \frac{b_3X^{III}}{\omega_c^3} + \frac{b_4Y^{IV}}{\omega_c^4} \cdots$$

$$+ c_0X + \frac{c_1Y^I}{\omega_c} - \frac{c_2X^{II}}{\omega_c^2} - \frac{c_3Y^{III}}{\omega_c^3} + \frac{c_4X^{IV}}{\omega_c^4} + \cdots \Bigg]$$

Finally, from eqn. (3), the distortion of the amplitude modulation of the emerging wave becomes

$$X\Bigg[b_0X + \frac{b_1Y^I}{\omega_c} - \frac{b_2X^{II}}{\omega_c^2} - \frac{b_3Y^{III}}{\omega_c^3} + \frac{b_4X^{IV}}{\omega_c^4} + \cdots$$

$$- c_0Y + \frac{c_1X^I}{\omega_c} + \frac{c_2Y^{II}}{\omega_c^2} - \frac{c_3X^{III}}{\omega_c^3} - \frac{c_4Y^{IV}}{\omega_c^4} + \cdots \Bigg]$$

$$+ Y\Bigg[b_0Y - \frac{b_1X^I}{\omega_c} - \frac{b_2Y^{II}}{\omega_c^2} + \frac{b_3X^{III}}{\omega_c^3} + \frac{b_4Y^{IV}}{\omega_c^4} - \cdots$$

$$+ c_0X + \frac{c_1Y^I}{\omega_c} - \frac{c_2X^{II}}{\omega_c^2} - \frac{c_3Y^{III}}{\omega_c^3} + \frac{c_4X^{IV}}{\omega_c^4} + \cdots \Bigg]$$

$$= b_0(X^2 + Y^2) + \frac{b_1}{\omega_c}(XY^I - X^IY) - \frac{b_2}{\omega_c^2}(XX^{II} + YY^{II})$$

$$- \frac{b_3}{\omega_c^3}(XY^{III} - X^{III}Y) + \cdots + \frac{c_1}{\omega_c}(XX^I + YY^I)$$

$$+ \frac{c_2}{\omega_c^2}(XY^{II} - X^{II}Y) - \frac{c_3}{\omega_c^3}(XX^{III} + YY^{III}) - \cdots$$

This is eqn. (4).

(8.2) Derivation of the Expressions for Intermodulation Distortion using Noise-Band Modulation

Lewin[6] has derived some very general results for the distortion due to powers of a flat noise band.

Suppose that the noise band has the form

$$\omega_M = a\sum_{\omega_m=1}^{\hat{\omega}_m} \cos(\omega_m t + \phi_{\omega_m}) \qquad . \quad . \quad . \quad (23)$$

Lewin's formulae give the amplitudes of distortion components, generated when ω_M is raised to some power, which are either coherent with components generated by some lower power of ω_M or incoherent with distortion due to all lower powers. The formulae are as follows:

(a) $(\omega_M)^\beta$ is contained in $(\omega_M)^\gamma$ ($\gamma > \beta$, and both are even or odd integers) with multiplying factor

$$\left(\frac{a\sqrt{\hat{\omega}_m}}{2}\right)^{\gamma-\beta} \frac{\gamma!}{\beta![(\gamma-\beta)/2]!} \qquad . \quad . \quad . \quad (24)$$

(b) The amplitude of the component of $(\omega_M)^\gamma$ (incoherent with terms arising from lower powers) having frequency $\omega(\omega \leqslant \hat{\omega}_m)$ is

$$\frac{a^\gamma}{2^{\gamma-1}}\Big[\gamma\{(\gamma\hat{\omega}_m - \omega)^{\gamma-1} - \gamma C_1[(\gamma-2)\hat{\omega}_m - \omega]^{\gamma-1}$$

$$+ \gamma C_2[(\gamma-4)\hat{\omega}_m - \omega]^{\gamma-1} \cdots\}\Big]^{1/2} \quad (25)$$

the series terminating when the coefficient of $\hat{\omega}_m$ is less than unity.

Expressions (24) and (25) can be used to deal with terms of the form $(\omega_M)^n$ and $(\omega_M)^{n-1}\omega_M'$ (the latter being obtained by differentiation), which constitute the leading terms in eqn. (5). In the absence of more extensive general formulae it seems necessary to treat the remaining terms in eqn. (5) individually. As an example, consider the third-order distortion due to the amplitude

characteristic. From eqns. (5) and (24) the third-order distortion is seen to be

$$\frac{b_5}{\omega_c^5}(K\omega_M^3 - 15\omega_M\omega_M'^2 - 10\omega_M^2\omega_M'')$$

where

$$K = \frac{b_3}{b_5}\omega_c^2 + 5a^2\hat{\omega}_m$$

Using eqn. (23), this becomes

$$\frac{b_5}{\omega_c^5}\Bigg[K\sum_{\omega_r} a\cos(\omega_r t + \phi_{\omega_r})\sum_{\omega_s} a\cos(\omega_s t + \phi_{\omega_s})\sum_{\omega_t} a\cos(\omega_t t + \phi_{\omega_t})$$

$$- 15\sum_{\omega_r} a\cos(\omega_r t + \phi_{\omega_r})\sum_{\omega_s}\omega_s a\sin(\omega_s t + \phi_{\omega_s})$$

$$\sum_{\omega_t}\omega_t a\sin(\omega_t t + \phi_{\omega_t}) + 10\sum_{\omega_r} a\cos(\omega_r t + \phi_{\omega_r})$$

$$\sum_{\omega_s} a\cos(\omega_s t + \phi_{\omega_s})\sum_{\omega_t}\omega_t^2 a\cos(\omega_t t + \phi_{\omega_t}) \Bigg]$$

where the summations extend from 1 to $\hat{\omega}_m$.

The term having frequency $\omega_r + \omega_s + \omega_t$ is

$$\tfrac{1}{4}a^3\frac{b_5}{\omega_c^5}\Bigg\{\sum_{\omega_r}\sum_{\omega_s}\sum_{\omega_t}(K + 15\omega_s\omega_t$$

$$+ 10\omega_t^2)\cos[(\omega_r + \omega_s + \omega_t)t + \phi_{\omega_r} + \phi_{\omega_s} + \phi_{\omega_t}]$$

First group together all terms which have the same phase angle, and hence add voltage-wise. Suppose that ω_r, ω_s and ω_t have specific values given by $\omega_r = \omega_x$, $\omega_s = \omega_y$, $\omega_t = \omega_z$ where $\omega_x > \omega_y > \omega_z$, the inequalities ensuring that all terms having frequencies $\omega_x + \omega_y + \omega_z$ will be incoherent.

Then the terms excluded are given by:

(a) $\omega_r = \omega_x$ $\omega_s = \omega_z$ $\omega_t = \omega_y$
(b) $\omega_r = \omega_x$ $\omega_s = \omega_x$ $\omega_t = \omega_z$
(c) $\omega_r = \omega_y$ $\omega_s = \omega_z$ $\omega_t = \omega_x$
(d) $\omega_r = \omega_z$ $\omega_s = \omega_x$ $\omega_t = \omega_y$
(e) $\omega_r = \omega_z$ $\omega_s = \omega_y$ $\omega_t = \omega_x$

Thus the amplitude of the term under consideration becomes

$$K + 15\omega_y\omega_z + 10\omega_z^2 + K + 15\omega_z\omega_y + 10\omega_y^2 + K$$

$$+ 15\omega_x\omega_z + 10\omega_z^2 + K + 15\omega_z\omega_x + 10\omega_x^2$$

$$+ K + 15\omega_x\omega_y + 10\omega_y^2 + K + 15\omega_y\omega_x + 10\omega_x^2$$

$$= 6K + 30(\omega_x\omega_y + \omega_y\omega_z + \omega_z\omega_x) + 20(\omega_x^2 + \omega_y^2 + \omega_z^2)$$

The term having frequency $\omega_r + \omega_s + \omega_t$ can therefore be written

$$\tfrac{1}{4}a^3\frac{b_5}{\omega_c^5}\Bigg\{\sum_{\omega_r=3}^{\hat{\omega}_m}\sum_{\omega_s=2}^{\omega_r}\sum_{\omega_t=1}^{\omega_t}[6K + 30(\omega_r\omega_s + \omega_s\omega_t + \omega_t\omega_r)$$

$$+ 20(\omega_r^2 + \omega_s^2 + \omega_t^2)]\times\cos[(\omega_r + \omega_s$$

$$+ \omega_t)t + \phi_{\omega_r} + \phi_{\omega_s} + \phi_{\omega_t}]$$

Putting $\omega_r + \omega_s + \omega_t = \omega_m$ and rearranging slightly, this becomes

$$\tfrac{1}{4}a^3\frac{b_5}{\omega_c^5}\Bigg(\sum_{\omega_r=3}^{\hat{\omega}_m}\sum_{\omega_s=2}^{\omega_r}\{6K + 10[2\omega_m^2 - \omega_m(\omega_r + \omega_s) + \omega_r^2$$

$$+ \omega_s^2 + \omega_r\omega_s]\}\times\cos(\omega_m t + \phi_{\omega_r} + \phi_{\omega_s} + \phi_{\omega_m - \omega_r - \omega_s}$$

Since these terms are randomly phased they will add power-wise

hus, writing the summations as integrals, the total contribution to the distortion term of frequency ω may be written as

$$a^3 \frac{b_5}{\omega_c^5} \left(\int_{\frac{1}{2}\omega}^{\frac{1}{2}\omega} \int_{\frac{1}{2}(\omega-\omega_r)}^{\omega_r} \left\{ 6K + 10[2\omega^2 - \omega(\omega_r + \omega_s) + \omega_r^2 + \omega_s^2 + \omega_r\omega_s] \right\}^2 d\omega_s d\omega_r + \int_{\frac{1}{2}\omega}^{\omega} \int_{\frac{1}{2}(\omega-\omega_r)}^{\omega-\omega_r} \left\{ 6K + 10[2\omega^2 - \omega(\omega_r + \omega_s) \right. \right.$$

$$\left. \left. + \omega_r^2 + \omega_s^2 + \omega_r\omega_s] \right\}^2 d\omega_s d\omega_r \right)^{1/2} . \quad (26)$$

The limits of these double integrals are arrived at as follows. ..nce $\omega_r + \omega_s + \omega_t = \omega$, $\omega_r > \omega_s > \omega_t$ and $\omega_t > 0$, we must ..ave $\frac{1}{3}\omega < \omega_r < \omega$. Also, from these conditions it follows that $.(\omega - \omega_r) < \omega_s < \omega_r$ (since ω_s will have its least value when .. $= \omega_s$). It is apparent that the integration must be split into ..o portions, since ω_s can reach its maximum permitted value, ... ω_r, only when $\omega_r < \frac{1}{2}\omega$. One portion is evidently obtained ... allowing ω_r to lie in the range $\frac{1}{3}\omega$–$\frac{1}{2}\omega$ and ω_s to lie in the ..nge $\frac{1}{2}(\omega - \omega_r)$–$\omega_r$, and this is the first integral of expres-..n (26). When $\omega_r > \frac{1}{2}\omega$, the maximum possible value of ω_s is

given by $\omega_s = \omega - \omega_r$, and hence we have the second integral of expression (26).

Four more similar double integrals are arrived at by putting successively $\omega_r + \omega_s - \omega_t = \omega$, $\omega_r - \omega_s + \omega_t = \omega$, $\omega_r - \omega_s - \omega_t = \omega$ and $\omega_r - \omega_s - \omega_t = -\omega$. The required result is obtained by evaluating all six integrals and adding the separate contributions on a power basis. This evaluation requires considerable algebraic manipulation, owing to the number of terms generated by the initial squaring and by the forms of the limits of the inner integrals. The final expression involves the square root of a polynomial of sixth degree in $\hat{\omega}_m$ and ω which, when $\omega = \hat{\omega}_m$, reduces to the second entry in Table 1.

DESIGN DATA FOR SYMMETRICAL DARLINGTON FILTERS

By J. K. SKWIRZYNSKI, B.Sc., and J. ZDUNEK, Dipl.Ing.

SUMMARY

The aim of the paper is to produce complete design data (i.e. the values of components in terms of generalized frequency and load conditions) of symmetrical equally terminated filters consisting of reactances only. The data are produced in graphical form and are carefully plotted to slide-rule accuracy; they are sufficiently dense to cover fully the usual requirements, without the use of interpolation.

The filters are designed on the basis of Darlington's insertion-loss theory with Chebȳshev behaviour both in the pass band and in the attenuation band. This allows for the unique correlation of the usual design specifications, which is displayed in a graphical form.

The underlying theory of the quadripoles and of the realization procedure is explained fully and independently of the practical part of the paper.

Finally, proper alignment procedure is described.

LIST OF SYMBOLS

I = General current.
Z = Impedance.
A, B, C, D = Chain matrix elements.
R = Termination resistances.
P = Power.
η = Efficiency (power ratio).
V = Complex voltage = $V\varepsilon^{j\phi}$.
$\Lambda(p) = V_{2o}/V_2$ = Voltage insertion-loss function (complex).
A, A', B, B' = Polynomials in $p^2 = -\Omega^2$ entering into $\Lambda(p)$.
$p = j\Omega$.
Ω = Generalized frequency.
α_s = Real part of the sth zero of $\Lambda(p)$.
ρ_s = Modulus of the sth zero of $\Lambda(p)$.
$p_{\infty s}$ = sth pole of $\Lambda(p)$.
J, J' = Arbitrary constants ($J^2 = J'^2$).
$\Xi(p)$ = Auxiliary function, derived from $\Lambda(p)$.
φ = Auxiliary function.
A_p = Maximum attenuation in the pass band.
A_a = Minimum attenuation in the suppression band.
k = Bandwidth parameter.
K = Complete elliptic integral.
$q(\theta)$ = Elliptic modular function.
sn = Jacobi's elliptic function.
P = Denominator of $\Lambda(p)$.
P^+, P^-, B^+, B^- = Polynomials in $p^2 = -\Omega^2$ entering into $\Xi(p)$.
n = Number of sections.
a_s = Ladder coefficients.
f = Frequency, c/s.
B_a, B_p = Bandwidths corresponding to A_a and A_p, respectively.
f_0 = Mid-band frequency.
$\omega = 2\pi f$.
ω_B = Characteristic angular frequency of the filter.

L = Inductance.
C = Capacitance.
$\Omega_{\infty s,1,2}$ = Normalized frequency in band-pass filters.
Ω_{0s} = sth zero of $\varphi(\Omega)$.
$\Omega_{\infty s} = \dfrac{1}{\Omega_{0s}}$ = sth pole of $\varphi(\Omega)$.
Γ = Reflection factor.

INTRODUCTION

The specifications of the filters required in modern communication systems are often so stringent that they cannot easily be met by normal image-parameter theory.

In 1931–32 Norton[11] and Bennett[3] gave an exact solution for the case of maximally flat behaviour in the pass band with finite zeros in the attenuation band. This was followed 20 years later by solutions for Chebȳshev behaviour in the pass band by Dishal,[7] Belevitch,[2] Orchard[12] and Green.[8] These effectively replaced the image-parameter theory of 'constant-k' filters by an exact 'insertion loss' (or some other analogous quantity) theory, but gave no help for the more important 'm-derived' types of filters. However, in 1939, Darlington, in an epoch-making paper,[6] gave a general method of solution for this and allied problems. Other papers on this subject have been published by Cauer,[4] Norton,[11] Cocci,[5] Piloty,[13] Bader,[1] *et al.* All these methods lead to tedious computations, and thus have remained in the use of the network specialist rather than the practising engineer.

The aim of the paper is to provide a solution for a small but important class of networks that can be used by the practical engineer rather than the network specialist. These are the so-called symmetrical and equally terminated filters consisting of one or two sections (i.e. three or five branches) having Chebȳshev behaviour in both the pass and the attenuation bands. A typical characteristic of a two-section low-pass filter is shown in Fig. 1. Filters with two sections will satisfy most of the commonly met
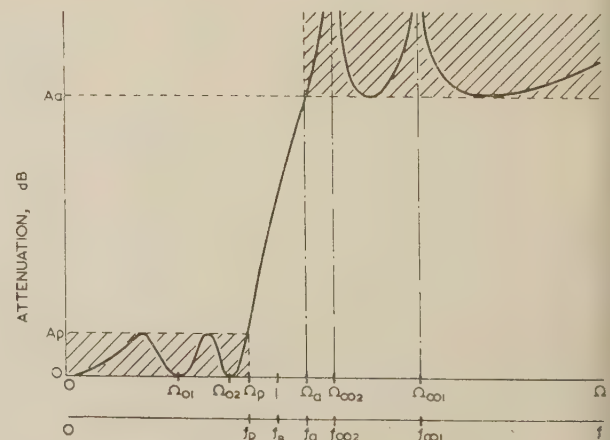
Fig. 1.—The insertion-loss function for a low-pass filter (two sections) with Chebȳshev pass- and attenuation-band behaviour.

requirements, while, on the other hand, symmetrical purely reactive filters were chosen because, for equal resistive terminations, their normalized low-pass components can be explicitly expressed in terms of the insertion-loss function parameters. The actual filter elements (capacitances and inductances) for low-pass, high-pass, band-pass and band-stop filters can be easily computed (with slide-rule accuracy) from the data given in Tables and formulae listed in proper Sections of the paper.

The paper is divided into three parts. In Part 1, some background is presented, explaining the insertion-loss theory and the Darlington method of synthesis of symmetrical filters to those who are interested in the method on which the design data are based.

Part 2 contains the Tables and graphs and the explanation of their use; it can be used directly without reference to Part 1.

Part 3 gives practical hints for alignment of the final network.

Finally, in the Appendix, an example is worked out for a low-pass filter, showing the use of the Tables and graphs.

Part 1. THEORETICAL INTRODUCTION

No new theoretical principles are introduced in this Part; it is hoped, however, that it will serve a useful purpose by explaining on a broad basis the method of synthesis of a symmetrical equally terminated filter.

The well-known quadripole parameters, such as impedance, reflection factor, etc., are related to the components of a general power insertion-loss function in Sections 1 and 2. These relations are further simplified in Section 3 for the case of equal terminations.

Algebraic properties of an insertion-loss function in the complex-frequency plane and the special type of such a function applied to symmetrical networks, are discussed in Sections 4 and 5, respectively; in particular, it is shown how to construct such a rational function from the suitably correlated specification data.

The method of synthesis of an equally terminated symmetrical (lattice) network is discussed in Section 6, where the lattice impedances are identified in terms of the insertion-loss function components.

Finally, in Section 7, the resultant lattice is transformed to an equivalent π- or T-ladder structure, the normalized network coefficients of which are explicitly expressed in terms of the singularities (poles and zeros) of the insertion-loss function.

(1) QUADRIPOLE IMPEDANCES AND RELATED FUNCTIONS

The behaviour of a passive quadripole is best expressed in terms of four related quantities, as shown in Fig. 2, namely the
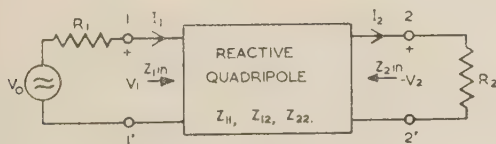
Fig. 2.—Terminated reactive quadripole.

input and output currents and voltages. The voltages can thus be obtained from currents in terms of the impedance matrix*

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & -Z_{12} \\ -Z_{12} & Z_{22} \end{bmatrix} \cdot \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} . \quad . \quad . \quad . \quad (1)$$

* It should be noted that all the loop currents are arrowed in the clockwise direction.

which contains only three independent parameters. Conversely:

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} Z_{11}/\Delta & Z_{12}/\Delta \\ Z_{12}/\Delta & Z_{22}/\Delta \end{bmatrix} \cdot \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \quad . \quad . \quad . \quad (2)$$

where

$$\Delta = Z_{11}Z_{22} - Z_{12}^2 \quad . \quad . \quad . \quad . \quad (3)$$

is the determinant of the Z matrix in eqn. (1).

Combining eqns. (1) and (2) we can obtain an alternative pair of relations between the input and the output currents and voltages in terms of the so-called 'chain matrix',

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} -V_2 \\ I_2 \end{bmatrix} \quad . \quad . \quad . \quad (4)$$

where

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} Z_{11}/Z_{12} & \Delta/Z_{12} \\ 1/Z_{12} & Z_{22}/Z_{12} \end{bmatrix} \quad . \quad . \quad . \quad (5)$$

Since the determinant of this matrix is clearly unity, eqn. (4) can be written conversely:

$$\begin{bmatrix} -V_2 \\ I_2 \end{bmatrix} = \begin{bmatrix} D & -B \\ -C & A \end{bmatrix} \cdot \begin{bmatrix} V_1 \\ I_1 \end{bmatrix} \quad . \quad . \quad . \quad (6)$$

From the above equation one can easily derive the well-known open- and short-circuit driving-point and transfer impedances in terms of the three independent network functions:

$$\left. \begin{aligned} \frac{V_1}{I_1}\Big|_{I_2=0} &= Z_{1,oc} = \frac{A}{C} = Z_{11} \\ \frac{V_2}{I_1}\Big|_{I_2=0} &= Z_{12,oc} = \frac{V_1}{I_2}\Big|_{I_1=0} = \frac{1}{-C} = -Z_{12} \\ \frac{V_2}{I_2}\Big|_{I_1=0} &= Z_{2,oc} = \frac{D}{C} = Z_{22} \\ \frac{V_1}{I_1}\Big|_{V_2=0} &= Z_{1,sc} = \frac{B}{D} = \frac{\Delta}{Z_{22}} \\ \frac{V_2}{I_1}\Big|_{V_1=0} &= Z_{12,sc} = \frac{V_1}{I_2}\Big|_{V_2=0} = B = \frac{\Delta}{Z_{12}} \\ \frac{V_2}{I_2}\Big|_{V_1=0} &= Z_{2,sc} = \frac{B}{A} = \frac{\Delta}{Z_{11}} \end{aligned} \right\} \quad . \quad (7)$$

(2) THE INSERTION-LOSS FUNCTION, REFLECTION FACTORS, ETC.

Consider a purely reactive quadripole terminated at both ends with resistances R_1 and R_2, respectively, as shown in Fig. 2. A relation analogous to eqn. (1) can now be written:

$$\begin{bmatrix} V_0 \\ 0 \end{bmatrix} = \begin{bmatrix} R_1 + Z_{11} & -Z_{12} \\ -Z_{12} & R_2 + Z_{22} \end{bmatrix} \cdot \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} \quad . \quad (8)$$

while the output voltage becomes

$$V_2 = I_2 R_2 \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Removing the reactive quadripole (i.e. connecting the terminals 1–1′ and 2–2′ directly), the output voltage is

$$V_{2o} = V_o \frac{R_2}{(R_1 + R_2)} \quad . \quad . \quad . \quad . \quad (10)$$

while the corresponding power delivered to the load R_2 becomes

$$P_{2o} = \frac{V_{2o}^2}{R_2} = V_o^2 \frac{R_2}{(R_1 + R_2)^2} . \quad . \quad . \quad (11)$$

This power attains its maximum when $R_1 = R_2$:

$$P_{2o\,max} = \frac{V_o^2}{4R_1} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

We can thus define the 'loaded generator efficiency' as

$$\frac{P_{2o}}{P_{2o\,max}} = \eta_G = \frac{4R_1R_2}{(R_1 + R_2)^2} \quad . \quad . \quad . \quad (13)$$

Reinserting the quadripole, as shown in Fig. 2, the power dissipated in the load becomes

$$P_2 = \frac{V_2^2}{R_2} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

so that the total 'system efficiency' is

$$\frac{P_2}{P_{2o}} = \eta \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Hence, using expression (8), we define the insertion power ratio as

$$\frac{P_{2o}}{P_2} = \left(\frac{V_{2o}}{V_2}\right)^2 = \frac{1}{\eta} = \Lambda^2(p) = \Lambda(p)\Lambda(-p) \quad . \quad (16)$$

where, from eqns. (8), (9) and (10),

$$\frac{V_{2o}}{V_2} = \frac{(R_1Z_{22} + R_2Z_{11}) + (R_1R_2 + Z_{11}Z_{22} - Z_{12}^2)}{(R_1 + R_2)Z_{12}} \quad . \quad (17)$$

For a purely reactive quadripole the first term in the numerator is imaginary and the second is real.

Thus $$\Lambda(p) = \frac{A + pB}{P} \quad . \quad . \quad . \quad . \quad (17a)$$

where, A, B and P are even polynomials in p, the complex generalized frequency, and

$$\Lambda^2(p) = \frac{(R_1Z_{22} + R_2Z_{11})^2 - (R_1R_2 + Z_{11}Z_{22} - Z_{12}^2)^2}{(R_1 + R_2)^2 Z_{12}^2}$$

$$= \frac{A^2 - p^2B^2}{P^2} = \frac{N}{P^2} = \frac{(A + pB)}{P}\frac{(A - pB)}{P} \quad . \quad (18)$$

The insertion-loss power ratio, eqn. (16), can also be expressed in an alternative very useful form [obtained by adding and subtracting appropriate terms in eqn. (18)]:

$$\frac{P_{2o}}{P_2} = \eta_G + \frac{(R_1Z_{22} - R_2Z_{11})^2 - (R_1R_2 - Z_{11}Z_{22} + Z_{12}^2)^2}{(R_1 + R_2)^2Z_{12}^2}$$

$$= \eta_G + \frac{A'^2 - p^2B'^2}{P^2} = \eta_G + \frac{N'}{P^2}$$

$$= \eta_G + \frac{A' + pB'}{P}\frac{A' - pB'}{P} = \eta_G(1 + \phi^2) \quad . \quad . \quad (19)$$

where A' and B' are even polynomials in p analogous to A and B.
Eqns. (18) and (19) enable us to make the following identifications:

$$\frac{A}{P} = \frac{R_1Z_{22} + R_2Z_{11}}{(R_1 + R_2)Z_{12}}, \quad \frac{A'}{P} = \frac{R_1Z_{22} - R_2Z_{11}}{(R_1 + R_2)Z_{12}}$$
$$\frac{pB}{P} = \frac{R_1R_2 + Z_{11}Z_{22} - Z_{12}^2}{(R_1 + R_2)Z_{12}}, \quad \frac{pB'}{P} = \frac{R_1R_2 - Z_{11}Z_{22} + Z_{12}^2}{(R_1 + R_2)Z_{12}} \quad\Bigg\} \quad (20)$$

which in turn lead to relations between the general network parameters as defined in the preceding Section and the component parts of the insertion-loss function:

$$A = \frac{R_1 + R_2}{2R_2}\frac{A - A'}{P} \quad B = \frac{R_1 + R_2}{2}\frac{p(B - B')}{P}$$
$$C = \frac{R_1 + R_2}{2R_1R_2}\frac{p(B + B')}{P} \quad D = \frac{R_1 + R_2}{2R_1}\frac{A + A'}{P} \quad\Bigg\} \quad . \quad (2$$

Using the above and eqns. (7), we can correlate the short- and open-circuit driving-point and transfer impedances with the components of the power insertion-loss function:

$$Z_{1,oc} = R_1\frac{A - A'}{p(B + B')}$$
$$Z_{12,oc} = -\frac{2R_1R_2}{R_1 + R_2}\frac{P}{p(B + B')}$$
$$Z_{2,oc} = R_2\frac{A + A'}{p(B + B')}$$
$$Z_{1,sc} = R_1\frac{p(B - B')}{A + A'} \quad\Bigg\} \quad . \quad . \quad (2$$
$$Z_{12,sc} = \frac{R_1 + R_2}{2}\frac{p(B - B')}{P}$$
$$Z_{2,sc} = R_2\frac{p(B - B')}{A - A'}$$

In order to obtain the reflection factors in terms of the power insertion-loss function components, we proceed as follows: from eqns. (21), the chain matrix defined in eqn. (4) becomes

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \frac{R_1 + R_2}{2R_1R_2P}\begin{bmatrix} R_1(A - A') & R_1R_2p(B - B') \\ p(B + B') & R_2(A + A') \end{bmatrix} \quad . \quad (2$$

whose determinant is

$$1 = \frac{(R_1 + R_2)^2}{4R_1R_2P^2}\begin{vmatrix} A - A' & p(B - B') \\ p(B + B') & A + A' \end{vmatrix}$$

$$= \frac{1}{\eta_G P^2}(N - N') \quad . \quad . \quad . \quad . \quad . \quad (24$$

so that $$N' = N - \eta_G P^2 = N(1 - \eta_G\eta) = N\Gamma^2 \quad . \quad . \quad (25$$

where $$\Gamma^2 = \Gamma_1^2 = \Gamma_2^2 = \frac{P_{2o\,max} - P_2}{P_{2o\,max}} \quad . \quad . \quad (26$$

Γ_1 and Γ_2 being the moduli of the reflection factors as seen from the terminals 1–1' and 2–2', respectively, in Fig. 2. From eqns. (25) and (26),

$$\Gamma^2 = \frac{N'}{N} = \frac{A'^2 - p^2B'^2}{A^2 - p^2B^2} \quad . \quad . \quad . \quad (27$$

The individual reflection coefficients can now be shown to be

$$\Gamma_1 = \frac{R_1 - Z_{1in}}{R_1 + Z_{1in}} = \frac{A' + pB'}{A + pB}$$
$$\Gamma_2 = \frac{R_2 - Z_{2in}}{R_2 + Z_{2in}} = \frac{-A' + pB'}{A + pB} \quad\Bigg\} \quad . \quad . \quad (28$$

where Z_{1in} and Z_{2in} are the input impedances of the terminated quadripole, as indicated in Fig. 2.
This gives a complete correlation between the various network parameters necessary for further synthesis.

(3) SYMMETRICAL AND EQUALLY TERMINATED QUADRIPOLES

A four-terminal network is said to be symmetrical with respect to its ends 1 and 2 [terminals 1–1′ and 2–2′ of present Fig. 2] when an interchange of these pairs of terminals has no effect upon the external behaviour.*

This requires that, in eqns. (1) and (4),

$$Z_{11} = Z_{22}; \quad A = D \quad . \quad . \quad . \quad . \quad (29)$$

so that a symmetrical quadripole depends on only two independent parameters, namely Z_{11} and Z_{12}.

The discussion in the paper will be confined to symmetrical quadripoles which are equally terminated, so that (see Fig. 2) $R_1 = R_2 = R$, and, as a consequence of eqns. (29) and (21), $C = 0$. Hence, the power ratio defined in eqn. (19) becomes

$$\left.\frac{P_{2o}}{P_2}\right|_{R_1 = R_2} = 1 - \frac{p^2 B'^2}{P^2} = 1 + \phi^2$$

$$= 1 - \left(\frac{R^2 - Z_{11}^2 + Z_{12}^2}{2RZ_{12}}\right)^2 \quad . \quad . \quad (30)$$

As mentioned in the Introduction, such a limitation enables us to determine completely the network parameters of a ladder structure which, for the low-pass case, can be realized either in T- or π-configuration [see Figs. 3(a) and 3(b)]. However, before

(a)

(b)
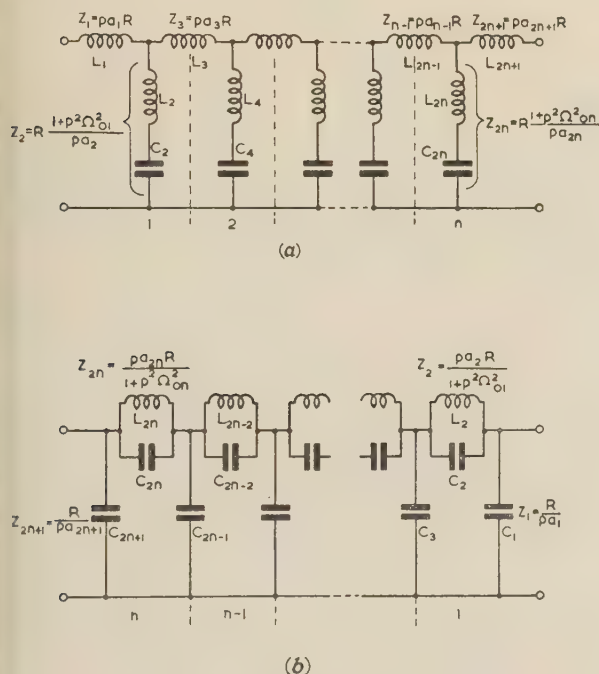
Fig. 3.—Basic configuration for low-pass filters.

(a) T-configuration.
(b) π-configuration.

proceeding with the actual determination of the network coefficients resulting in a required insertion-loss function, it will be useful to consider some of its general algebraic properties.

(4) SOME PROPERTIES OF THE INSERTION-LOSS FUNCTION

The insertion voltage ratio, as defined in eqn. (17), is a rational function of $p = j\Omega$, where Ω is a generalized normalized fre-

quency (see Section 8, where Ω is defined for every type of filter). Consider then a basic low-pass filter, such as that shown in Fig. 3(a) or 3(b), consisting of sections connected in ladder fashion, whose insertion voltage ratio can be written:

$$\frac{V_{2o}}{V_2} = \Lambda(p) = \frac{A + pB}{P} = J \frac{(p + \alpha_o)\prod\limits_{s=1}^{n}(p^2 + 2\alpha_s p + \rho_s^2)}{\prod\limits_{s=1}^{n}\left(1 - \frac{p^2}{p_{\infty s}^2}\right)} \quad . \quad (31)$$

where J is an arbitrary real constant. The even polynomials A and B are each of degree n in p^2. The roots of the denominator P are

$$\left.\begin{array}{c} p_{\infty s} \\ p_{\infty s}^* \end{array}\right\} = \pm j\Omega_{\infty s} = \Omega_{\infty s} \left/ \pm \frac{\pi}{2}\right. \quad . \quad . \quad . \quad (32)$$

where $\Omega_{\infty s}$ are the frequencies of infinite loss (to be called the 'poles' of the system). The numerator $A + pB$ has $2n + 1$ roots, of which

$$\left.\begin{array}{c} p_s \\ p_s^* \end{array}\right\} = -\alpha_s \pm j\sqrt{(\rho_s^2 - \alpha_s^2)} = \rho_s \left/ \pm \left(\frac{\pi}{2} + \phi_s\right)\right. . \quad (33)$$

are conjugate pairs with negative real parts, while

$$p_0 = -\alpha_0$$

is real. These roots are the natural modes of the system and will be called 'zeros'. Fig. 4(a) shows a typical zero (O) and
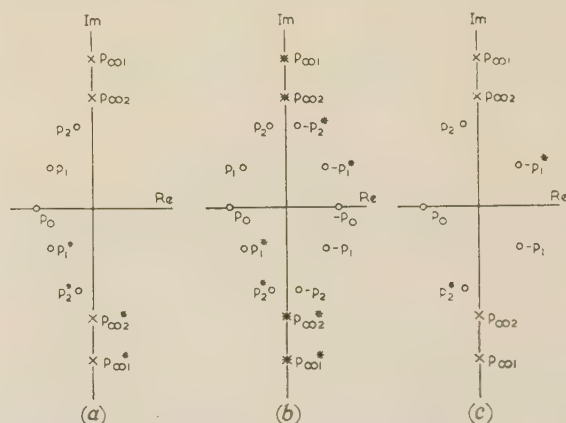
Fig. 4.—The distribution of poles and zeros in the complex frequency p-plane.

(a) Singularities of $\Lambda(p)$.
(b) Singularities of $\Lambda(p)\Lambda(-p)$.
(c) Singularities of Ξ.
The case illustrated corresponds to $n = 2$.

pole (X) configuration in the complex frequency plane for $n = 2$. The voltage insertion function as defined in eqn. (31) will thus have n finite positive frequencies of infinite loss, and furthermore, the loss will be infinite also when $\Omega \to \infty$; by suitably defining the arbitrary constant J [namely by making

$$J^{-1} = \alpha_0 \prod_{s=1}^{n} \rho_s^2$$

so that $\Lambda(0) = 1$], one can ensure zero loss when $\Omega = 0$.

The corresponding power insertion-loss function, eqn. (18), becomes

$$\frac{P_{2o}}{P_2} = \Lambda(p)\Lambda(-p) = J^2 \frac{(p_o^2 - p^2)\prod\limits_{s=1}^{n}(p^2 - p_s^2)(p^2 - p_s^{*2})}{\prod\limits_{s=1}^{n}\left(1 - \frac{p^2}{p_{\infty s}^2}\right)^2} \quad . \quad (34)$$

The power loss function is necessarily real; its zeros and poles are twice the number of those defined in eqn. (31); they are symmetrically situated on both sides of the real frequency (i.e. imaginary p) axis, as shown in Fig. 4(b), where the double poles are indicated by an asterisk.

It will be found convenient in Section 6 to construct another function $\Xi(p)$ which can be directly obtained from $\Lambda(p)$ defined above by changing the signs of every odd α_s (i.e. every α for odd s). Thus:

$$\Xi(p) = J' \frac{(p + \alpha_0) \prod_{s=1}^{n} [p^2 + 2(-1)^s \alpha_s p + \rho_s^2]}{P} \ ; \ J^2 = J'^2 \ . \ (35)$$

Here we have selected all odd conjugate zero pairs from $\Lambda(-p)$ as defined in eqn. (31), and the remaining even conjugate zero pairs from its conjugate $\Lambda(p)$. The zeros and poles of $\Xi(p)$ are shown in Fig. 4(c) for $n = 2$. It should be noted that

$$\Xi(p)\Xi(-p) = \Lambda(p)\Lambda(-p) = \Lambda^2(p) \ . \ . \ (36)$$

gives the real power insertion-loss function as defined in eqn. (34).

(5) THE CHOICE OF THE INSERTION-LOSS FUNCTION FOR A SYMMETRICAL EQUALLY TERMINATED NETWORK

For a symmetrical equally terminated network, the insertion-loss function defined in eqn. (30) depends on an odd function $\varphi(p)$. Thus, let

$$B' = J' \sum_{q=0}^{n} B_q' p^{2q} = J' \prod_{s=1}^{n} (p^2 + \Omega_{0s}'^2) \ . \ . \ (37)$$

and

$$\varphi = J' \frac{j\Omega \prod_{s=1}^{n} (\Omega^2 - \Omega_{0s}'^2)}{\prod_{s=1}^{n} \left(1 - \frac{\Omega^2}{\Omega_{\infty s}^2}\right)} \ . \ . \ . \ (38)$$

so that

$$P = \prod_{s=1}^{n} \left(1 - \frac{\Omega^2}{\Omega_{\infty s}^2}\right) \ . \ . \ . \ . \ (39)$$

The choice of this function will be governed by the filtering requirements of the network. Let us assume, for a basic low-pass filter, a certain limited distortion in the pass band not exceeding A_p decibels, while for the stop band we require a signal suppression not less than A_a decibels; furthermore, the change from the pass band to the attenuation band should be made arbitrarily narrow along the frequency axis. The general situation is shown in Fig. 1. It may be noted that, for a purely reactive and equally terminated filter, A_p gives also the discrimination in the pass band, and is thus related to the corresponding reflection factor Γ_p.

To simplify the analysis we shall adopt Cauer's[4] and Darlington's[6] choice by making the φ-function reciprocal about $\Omega = 1$, so that we need consider this function only in the range $0 \leqslant \Omega \leqslant 1$, thereby automatically defining it in the remaining range $1 \leqslant \Omega \leqslant \infty$ on the positive real frequency axis. Thus, let

$$\varphi(\Omega) = \frac{J^2}{\varphi(1/\Omega)} \ . \ . \ . \ . \ . \ (40)$$

The necessary and sufficient condition for this relation to hold is that, in eqn. (38),

$$\Omega_{0s}' = \Omega_{0s} = \frac{1}{\Omega_{\infty s}} \ . \ . \ . \ . \ (41)$$

Then the chosen φ-function becomes

$$\varphi(\Omega) = J' \frac{j\Omega \prod_{1}^{n} (\Omega^2 - \Omega_{0s}^2)}{\prod_{1}^{n} (1 - \Omega^2 \Omega_{0s}^2)} \ . \ . \ . \ (42)$$

A typical logarithmic insertion-loss function obtained from eqns. (42) and (30), i.e.

$$10 \log_{10} \frac{P_{2o}}{P_2}$$

is shown in Fig. 1, where the maximum attenuation in the pass band is given by

$$A_p = 10 \log_{10} \frac{P_{2o}}{P_2}\bigg|_{max} = -10 \log_{10}(1 - \Gamma_p^2) \text{ when } 0 < \Omega \leqslant \Omega$$

and the minimum attenuation in the suppression band is given by

$$A_a = 10 \log_{10} \frac{P_{2o}}{P_2}\bigg|_{min} \text{ when } \Omega_a \leqslant \Omega < \infty$$

Because of the reciprocal nature of the insertion-loss function we must have

$$\sqrt{(\Omega_p \Omega_a)} = 1 \ . \ . \ . \ . \ . \ (43)$$

Let, also,

$$\frac{\Omega_p}{\Omega_a} = k < 1 \ . \ . \ . \ . \ (44)$$

Then

$$\Omega_p = \sqrt{k}; \ \Omega_a = \frac{1}{\sqrt{k}} \ . \ . \ . \ . \ (45)$$

Following Cauer and Darlington we shall specify the position of the zeros and the poles of the φ-function in terms of Jacobi's elliptic functions. It can then be shown that the four quantities A_a, A_p, k and n [number of sections, as in Fig. 3(a) or 3(b)] are approximately correlated together as follows:

$$A_a \simeq 10 \log_{10} (10^{A_p/10} - 1) - 10 (2n + 1) \log_{10} q(\theta) - 12 \cdot 041$$
$$. \ . \ . \ . \ (46)$$

where $\theta = \arc \sin k$ and $q(\theta)$ is the elliptic modular function tabulated in Reference 9, pp. 49–51. This relation allows us to estimate the performance of the filter as explained in Section 9.

The zeros and poles of φ as defined in eqn. (38) are now given in terms of Jacobi's elliptic functions, namely

$$\Omega_{0s} = \sqrt{k} \ \text{sn} \left(\frac{2sK}{2n+1}, k\right) = \sqrt{k} \sin \psi_s \ . \ . \ (47)$$

where

$$\frac{2sK}{2n+1} = u_s = \int_0^{\psi_s} \frac{d\beta}{\sqrt{(1 - k^2 \sin^2 \beta)}} \ . \ . \ (48)$$

and K is the complete elliptic integral of the first kind:

$$K = K(k) = \int_0^{\pi/2} \frac{d\beta}{\sqrt{(1 - k^2 \sin^2 \beta)}} \ . \ . \ (49)$$

These functions are conveniently tabulated in Reference 9, p. 85 et seq., and in Reference 10.

Darlington,[6] following Norton,[11] has shown that, given the poles of an insertion-loss function of the type shown in Fig. 1, its zeros, i.e. the zeros of the numerator in eqn. (30) or (34), can be obtained in terms of the elliptic modular transformation, giving

$$\alpha_0 = \sqrt{k} \tan \psi_0 \ . \ . \ . \ . \ (50)$$

where
$$\sin \psi_0 = \text{sn} (u_0, k')$$
$$k' = \sqrt{(1 - k^2)}$$

and
$$u_0 \simeq \frac{2K}{(2n + 1)\pi} \text{ arc sinh } (10^{A_p/10} - 1)^{-1/2} . \quad (51)$$

Also
$$\left.\begin{array}{l} \alpha_s = \dfrac{\alpha_0 \sqrt{[1 - \Omega_{0s}^2(k + k^{-1} - \Omega_{0s}^2)]}}{1 + \alpha_0^2 \Omega_{0s}^2} \\[2mm] \rho_s^2 = \dfrac{\alpha_0^2 + \Omega_{0s}^2}{1 + \alpha_0^2 \Omega_{0s}^2} \end{array}\right\} \quad (52)$$

The quantities α_s and ρ_s are, respectively, the real parts and the moduli of the zeros in the complex p-plane as defined in eqns. (31)–(33). It should be noted here that α_0 and Ω_{0s} completely specify the filter response for given values of k and A_p [which in turn are related to n and A_a by eqn. (46)].

(6) SYNTHESIS OF AN EQUALLY TERMINATED SYMMETRICAL LATTICE

A lattice structure as shown in Fig. 5 can be designed by observing certain relations between functions derived from its
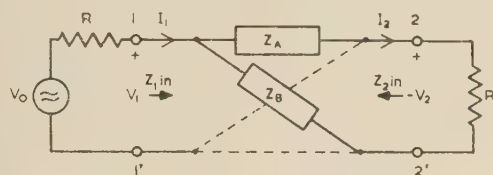
Fig. 5.—Symmetrical lattice.

insertion-loss expression. The two independent lattice impedances become

$$\left.\frac{V_2}{I_1}\right|_{I_2=0} = \tfrac{1}{2}(Z_A - Z_B) = -Z_{12} = Z_{12,oc} \quad . \quad (53)$$

$$\left.\frac{V_1}{I_1}\right|_{I_2=0} = \tfrac{1}{2}(Z_A + Z_B) = Z_{11} = Z_{22} = Z_{11,oc} \quad . \quad (54)$$

The insertion voltage ratio is, from eqn. (17),

$$\left.\frac{V_{2o}}{V_2}\right|_{R_1=R_2=R} = \frac{2RZ_{11} + R^2 + Z_{11}^2 - Z_{12}^2}{2RZ_{12}}$$
$$= \frac{(R + Z_A)(R + Z_B)}{R(Z_A - Z_B)} \quad . \quad (55)$$

But the insertion power-loss ratio, as defined in eqns. (30), (35) and (36), can also be written:

$$\frac{P_{2o}}{P_2} = \Xi(p)\Xi(-p) = 1 - \frac{p^2 B'^2}{P^2} \quad . \quad . \quad (56)$$

where
$$\Xi(p) = 1 + \frac{pB'}{P} \quad . \quad . \quad . \quad (57)$$

since both B' and P are even polynomials in p. From eqn. (30) we can identify

$$\Xi(p) = 1 + \frac{R^2 - Z_{11}^2 + Z_{12}^2}{2RZ_{12}} \quad . \quad . \quad (58)$$

which, on substituting from eqns. (53) and (54), becomes

$$\Xi(p) = \frac{(R + Z_A)(R - Z_B)}{R(Z_A - Z_B)} \quad . \quad . \quad (59)$$

It should be noted that this expression is not equal to the insertion voltage ratio as defined in eqn. (55), since here Z_B is replaced in the numerator (and only there) by its conjugate value. Expression (59) can be written more conveniently:

$$\Xi(p) = \frac{\left(1 + \dfrac{Z_A}{R}\right)\left(1 - \dfrac{R}{Z_B}\right)}{\left(1 - \dfrac{Z_A}{Z_B}\right)} \quad . \quad . \quad (60)$$

With the aid of eqn. (35) we can define

$$P^+ - pB^+ = J' \prod_{s=1}^{s \leqslant \frac{1}{2}(n+1)} (p^2 - 2\alpha_{2s-1}p + \rho_{2s-1}^2) . \quad (61)$$

which is that part of $\Xi(p)$ which has its roots on the right-hand side of the complex frequency plane. Similarly, let us define

$$P^- + pB^- = (p + \alpha_0) \prod_{s=1}^{s \leqslant \frac{1}{2}n} (p^2 + 2\alpha_{2s}p + \rho_{2s}^2) \quad . \quad (61a)$$

which is that part of $\Xi(p)$ which has its roots on the left-hand side of the complex frequency plane. Note that the last product factor is designated by an integer less than or equal to $n/2$ in eqn. (61a), and less than or equal to $(n + 1)/2$ in eqn. (61). Hence,

$$\Xi(p) = \frac{P + pB'}{P} = \frac{(P^- + pB^-)(P^+ - pB^+)}{P}$$
$$= \frac{\left(1 + \dfrac{pB^-}{P^-}\right)\left(1 - \dfrac{pB^+}{P^+}\right)}{1 - p^2 \dfrac{B^- B^+}{P^- P^+}} \quad . \quad . \quad . \quad (62)$$

since
$$P = P^+ P^- - p^2 B^+ B^-$$

Identifying eqn. (62) with eqn. (60), we arrive at the lattice impedances expressed directly in terms of the known parameters of the power insertion-loss function of the system:

$$Z_A = R\frac{pB^-}{P^-}; \quad Z_B = R\frac{P^+}{pB^+} \quad . \quad . \quad (63)$$

At each pole, i.e. when $p = p_{\infty s}$ so that $P = 0$, one gets $Z_A - Z_B = 0$.

Thus
$$Z_A|_{p=p_{\infty s}} = Z_B|_{p=p_{\infty s}} . \quad . \quad . \quad . \quad (64)$$

It should be added here that there exist four possibilities of defining Z_A or Z_B, namely by interchanging the definitions (63), or by replacing the impedances by their duals.

The constant parameters α_0, α_s, ρ_s^2 and Ω_{0s} of the polynomial $P + pB'$ are obtained from the initial network specifications as explained in Section 5; all these quantities are positive by definition.

It should be noted that once all the pole and zero locations in the complex frequency plane have been obtained one should check that

$$\alpha_0 (\prod_1^n \rho_s^2 \Omega_{0s}^2 - 1) = 2 \sum_1^n (-1)^s \alpha_s \quad . \quad . \quad (65)$$

This condition ensures that the resulting lattice will have an equivalent ladder structure with directly realizable parameters, i.e. without mutual inductances. Eqn. (65) is arrived at by comparing the design procedure described here with the more general one used by Darlington (Part II of Reference 6); its derivation, although simple, is too lengthy to be included here.

(7) EXAMPLE OF THE DERIVATION OF LADDER ELEMENT VALUES BY MEANS OF AN EQUIVALENT LATTICE

Consider a two-section low-pass filter whose voltage insertion ratio, eqn. (31), is given in terms of the parameters calculated as explained in Section 5:

$$\frac{V_{2o}}{V_2} = \Lambda(p) = J \frac{(p + \alpha_0) \prod_1^2 (p^2 + 2\alpha_s p + \rho_s^2)}{\prod_1^2 (1 + p^2 \Omega_{0s}^2)}$$

$$= \frac{A + pB'}{P} \quad . \quad . \quad . \quad . \quad . \quad (66)$$

It is required that $\Lambda(0) = 1$, so that

$$J^{-1} = \alpha_0 \rho_1^2 \rho_2^2 \quad . \quad . \quad . \quad . \quad . \quad (67)$$

The auxiliary function $\Xi(p)$ is now formed directly from eqn. (66), as explained in the preceding Section:

$$\Xi(p) = J' \frac{(p + \alpha_0)(p^2 - 2\alpha_1 p + \rho_1^2)(p^2 + 2\alpha_2 p + \rho_2^2)}{(1 + p^2 \Omega_{01}^2)(1 + p^2 \Omega_{02}^2)}$$

$$= \frac{P + pB'}{P} = \frac{(P^- + pB^-)(P^+ - pB^+)}{P^- P^+ - p^2 B^- B^+} \quad . \quad . \quad (68)$$

Hence:

$$\left. \begin{array}{l} P^- = p^2(\alpha_0 + 2\alpha_2) + \rho_2^2 \alpha_0 \\ B^- = p^2 + \rho_2^2 + 2\alpha_0\alpha_2 \\ P^+ = J'(p^2 + \rho_1^2) \\ B^+ = J' 2\alpha_1 \end{array} \right\} \quad . \quad . \quad . \quad (69)$$

The lattice impedances given in eqns. (63) can now be conveniently expressed in the form

$$\left. \begin{array}{l} Z_A = R \dfrac{pB^-}{P^-} = R \dfrac{p(\gamma_1 p^2 + \gamma_2)}{\beta_1 p^2 + 1} \\[2mm] Z_B = R \dfrac{P^+}{pB^+} = R \dfrac{\gamma_3 p^2 + \gamma_4}{p} \end{array} \right\} \quad . \quad . \quad (70)$$

where $\quad \gamma_1 = \dfrac{1}{\alpha_0 \rho_2^2}; \quad \gamma_2 = \dfrac{1}{\alpha_0} + \dfrac{2\alpha_2}{\rho_2^2}; \quad \gamma_3 = \dfrac{1}{2\alpha_1};$

$$\gamma_4 = \frac{\rho_1^2}{2\alpha_1}; \quad \beta_1 = \frac{\alpha_0 + 2\alpha_2}{\alpha_0 \rho_2^2} \quad . \quad . \quad . \quad (71)$$

Also, from eqn. (64),

$$Z_A|_{p=p_{\infty s}} = Z_B|_{p=p_{\infty s}} \quad . \quad . \quad . \quad (72)$$

where

$$p_{\infty s}^{-2} = -\Omega_{02}^2$$

Our aim is to express the ladder elements $a_1, a_2, \ldots a_5$ [see Figs. 3(a) or 3(b)] in terms of β_1 and γ_s in eqns. (71), and hence in terms of the known parameters in eqns. (69). Since the ladder and lattice structures in Figs. 3(a) or 3(b) and Fig. 5 are equivalent, we can identify their impedances at various frequencies. Thus, for a structure such as that shown in Fig. 3(a),

$$Z_{11}|_{p=p_{\infty 1}} = Z_B|_{p=p_{\infty 1}} = p_{\infty 1} R(\gamma_3 - \Omega_{01}^2 \gamma_4) \quad . \quad (73)$$

for the lattice; similarly, for the ladder,

$$Z_{11,T}|_{p=p_{\infty 1}} = a_1 p_{\infty 1} R \quad . \quad . \quad . \quad (74)$$

Hence $\quad a_1 = \gamma_3 - \Omega_{01}^2 \gamma_4 = \dfrac{1}{2\alpha_1}(1 - \rho_1^2 \Omega_{01}^2) \quad . \quad . \quad (75)$

Similarly, equating

$$Z_{22}|_{p=p_{\infty 2}} = Z_B|_{p=p_{\infty 2}} = Z_{22,T}|_{p=p_{\infty 2}}$$

we obtain $\quad a_5 = \gamma_3 - \Omega_{02}^2 \gamma_4 = \dfrac{1}{2\alpha_1}(1 - \rho_1^2 \Omega_{02}^2) \quad . \quad . \quad (76)$

The remaining parameters a_2, a_3 and a_4 can be obtained as follows:

$$Z_{11,T}\Big|_{p=p_{\infty 2}} = \left(Z_1 + \frac{Z_2 Z_3}{Z_2 + Z_3}\right)\Big|_{p=p_{\infty 2}} = Z_{22,T}\Big|_{p=p_{\infty 2}}$$

from which $\quad a_5 - a_1 = \dfrac{a_3 \delta}{\delta - a_2 a_3} \quad . \quad . \quad . \quad (77)$

where $\quad \delta = \Omega_{02}^2 - \Omega_{01}^2 \quad . \quad . \quad . \quad . \quad (78)$

Similarly, from

$$Z_{22,T}\Big|_{p=p_{\infty 1}} = \left(Z_5 + \frac{Z_3 + Z_4}{Z_3 Z_4}\right)\Big|_{p=p_{\infty 1}} = Z_{11,T}\Big|_{p=p_{\infty 1}}$$

we obtain $\quad a_5 - a_1 = -\dfrac{a_3 \delta}{\delta + a_3 a_4} \quad . \quad . \quad . \quad (79)$

Comparing eqns. (77) and (79),

$$\left. \begin{array}{l} a_2 = \dfrac{1}{\gamma_4} + \dfrac{\delta}{a_3} \\[2mm] a_4 = \dfrac{1}{\gamma_4} - \dfrac{\delta}{a_3} \end{array} \right\} \quad . \quad . \quad . \quad (80)$$

Finally, from the lattice,

$$-Z_{12,oc} = Z_{12} = \frac{Z_B - Z_A}{2} = R \frac{P^+ P^- - p^2 B^+ B^-}{2pB^+ P^-}$$

$$= R \frac{P}{p\left(\dfrac{2\beta_1}{\gamma_4}p^2 + \dfrac{2}{\gamma_4}\right)}$$

while the equivalent expression for the ladder becomes

$$Z_{12,T} = \frac{Z_2 Z_4}{Z_2 + Z_3 + Z_4}$$

$$= R \frac{P}{p[(a_2 a_3 a_4 + \Omega_{01}^2 a_4 + \Omega_{02}^2 a_2)p^2 + (a_2 + a_4)]}$$

Identifying corresponding terms in the denominators of both these expressions, we obtain

$$a_2 a_3 a_4 + \Omega_{01}^2 a_4 + \Omega_{02}^2 a_2 = \frac{2\beta_1}{\gamma_4} \quad . \quad . \quad (81)$$

and, substituting for a_2 and a_4 from eqns. (80), we have

$$a_3 = \gamma_4(2\beta_1 - \Omega_{01}^2 - \Omega_{02}^2)$$

$$= \frac{\rho_1^2}{2\alpha_1}\left[\frac{2(\alpha_0 + 2\alpha_2)}{\alpha_0 \rho_2^2} - \Omega_{01}^2 - \Omega_{02}^2\right] \quad . \quad (82)$$

Hence

$$\left. \begin{array}{l} a_2 = \dfrac{2}{a_3}(\beta_1 - \Omega_{01}^2) \\[2mm] a_4 = \dfrac{2}{a_3}(\beta_1 - \Omega_{02}^2) \end{array} \right\} \quad . \quad . \quad . \quad (83)$$

so that all the ladder parameters are determined in terms of the known co-ordinates of the poles and zeros in the complex frequency plane.

Similarly, for a single-section low-pass filter ($n = 1$),

$$\left. \begin{array}{l} a_1 = a_3 = \dfrac{1}{\alpha_0} \\[2mm] a_2 = \dfrac{4\alpha_1}{\rho_1^2} \end{array} \right\} \quad . \quad . \quad . \quad . \quad (84)$$

Part 2. PRACTICAL DESIGN DATA

As explained in the Introduction, the aim is to produce a very simple design procedure for symmetrical and equally terminated filters; the actual values of the normalized network coefficients [i.e. a_s, see Fig. 3(a) or 3(b)] were thus computed as functions of the quantities specifying the required insertion-loss function of the given filter. The results are plotted in graphical form so as to allow for the 3-figure accuracy which is quite sufficient in practical cases.

The Tables and graphs can be used only for standard filters which are either to pass or to reject definite bands of frequencies. The filters consist of standard π- or T-sections connected in ladder formation so as to form the so-called 'm-derived' configuration [see Figs. 3(a) and 3(b)]. The number of sections is at most two. Finally, one should emphasize once more that the load and generator resistances are to be equal.

The usual power-loss specifications for a filter, whether of low-pass, high-pass, band-pass or band-stop type, are: the effective pass band, say B_p cycles per second; the effective attenuation or suppression band, say B_a cycles per second; the minimum permissible attenuation in the suppression band, say A_a decibels; and the maximum permissible variation in the pass band, say A_p decibels. The ratio B_p/B_a or B_a/B_p (depending on the type of the filter) will be called the bandwidth ratio k.

Since the normalized network coefficients a_s are plotted in terms of k, whose meaning is not the same for all four types of filters, we shall first relate k, A_a and A_p for each filter type, whatever the number of sections n, and also indicate how to obtain the actual values of the reactances in terms of a_s; this will be done in Section 8.

As indicated in the Introduction and fully discussed in Section 5, the shape of the power insertion-loss function is carefully specified so as to correlate the values of k, A_a and A_p for a given value of n. This correlation and its bearing on the actual design procedure will be discussed in Section 9.

Finally, in Section 10, the graphs of the normalized network coefficients a_s will be fully described and explained.

(8) NETWORK PARAMETERS FOR VARIOUS TYPES OF FILTERS

We shall consider each of the four types separately.

(a) Low-Pass Filter (Fig. 6).

Power is to be passed with at most A_p decibels variation, within the range $0 \leqslant f \leqslant f_p$.

Power is to be suppressed by at least A_a decibels, within the range $f_a \leqslant f \leqslant \infty$.

Here

$$\left.\begin{array}{c} f_p = B_p \\ f_a = B_a \\ k = \dfrac{B_p}{B_a} \end{array}\right\} \quad \cdots \cdots \quad (85)$$

The normalized frequency for a low-pass filter is

$$\Omega = \frac{\omega}{\omega_B}; \quad \omega_B = 2\pi\sqrt{(B_a B_p)} \quad \cdots \quad (86)$$

Typical π- and T-sections are shown in Fig. 6, where the impedances of particular branches are expressed in terms of $p \ (= j\Omega)$, Ω_{0s} and the generalized ladder coefficients a_s; the actual network inductances L and capacitances C are given by the following equations, where R is the generator or load resistance.

For a π-section:

$$\left.\begin{array}{c} C_{(2s\mp1)\pi} = \dfrac{a_{2s\mp1}}{\omega_B R} \\[2mm] C_{(2s)\pi} = \dfrac{\Omega_{0s}^2}{\omega_B R a_{2s}} \\[2mm] L_{(2s)\pi} = \dfrac{R a_{2s}}{\omega_B} \end{array}\right\} \quad \cdots \cdots \quad (87)$$

For a T-section:

$$\left.\begin{array}{c} L_{(2s\mp1)T} = \dfrac{R a_{2s\mp1}}{\omega_B} \\[2mm] L_{(2s)T} = \dfrac{R\Omega_{0s}^2}{\omega_B a_{2s}} \\[2mm] C_{(2s)T} = \dfrac{a_{2s}}{\omega_B R} \end{array}\right\} \quad \cdots \cdots \quad (88)$$

(b) High-Pass Filter (Fig. 7).

Power is to be suppressed by at least A_a decibels within the range $0 \leqslant f \leqslant f_a$.

Power is to be passed with at most A_p decibels variation, within the range $f_p \leqslant f < \infty$.

Here

$$\left.\begin{array}{c} f_p = B_p \\ f_a = B_a \\ k = \dfrac{B_a}{B_p} \end{array}\right\} \quad \cdots \cdots \quad (89)$$

Fig. 6.—Typical low-pass filter sections (π- and T- configurations) and insertion-loss response.
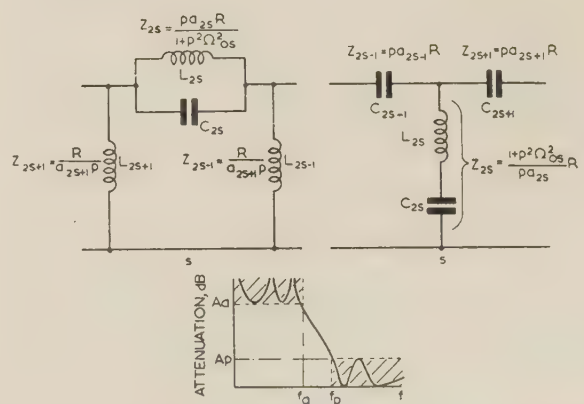
Fig. 7.—Typical high-pass filter sections (π- and T-configurations) and insertion-loss response.

The normalized frequency for a high-pass filter is

$$\Omega = -\frac{\omega_B}{\omega}; \quad \omega_B = 2\pi\sqrt{(B_aB_p)} \quad . \quad . \quad . \quad (90)$$

Typical π- and T-sections are shown in Fig. 7, and the values of the actual network components are given by the following equations.

For a π-section:

$$\left.\begin{aligned} L_{(2s\mp1)\pi} &= \frac{R}{\omega_B a_{2s\mp1}} \\[2mm] L_{(2s)\pi} &= \frac{Ra_{2s}}{\omega_B\Omega_{0s}^2} \\[2mm] C_{(2s)\pi} &= \frac{1}{\omega_B Ra_{2s}} \end{aligned}\right\} \quad . \quad . \quad . \quad (91)$$

For a T-section:

$$\left.\begin{aligned} C_{(2s\mp1)T} &= \frac{1}{\omega_B Ra_{2s\mp1}} \\[2mm] C_{(2s)T} &= \frac{a_{2s}}{\omega_B R\Omega_{0s}^2} \\[2mm] L_{(2s)T} &= \frac{R}{\omega_B a_{2s}} \end{aligned}\right\} \quad . \quad . \quad . \quad (92)$$

(c) Band-Pass Filter (Fig. 8).

Power is to be passed with at most A_p decibels variation, within the range $f_{p1} \leqslant f \leqslant f_{p2}$.
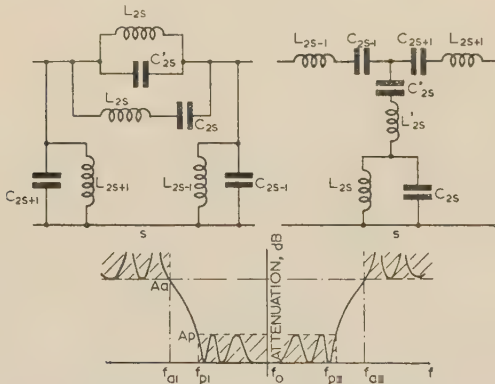
Fig. 8.—Typical band-pass filter sections (π- and T-configurations) and insertion-loss response.

Power is to be suppressed by at least A_a decibels, within the ranges $0 \leqslant f \leqslant f_{a1}$ and $f_{a2} \leqslant f < \infty$.

Here

$$\left.\begin{aligned} f_{p2} - f_{p1} &= B_p \\ f_{a2} - f_{a1} &= B_a \\ k &= \frac{B_p}{B_a} \end{aligned}\right\} \quad . \quad . \quad . \quad (93)$$

In order to preserve geometrical band-pass symmetry we have to correlate the band-edge frequencies with the mid-band frequency, f_0:

$$f_0^2 = f_{p1}f_{p2} = f_{a1}f_{a2} \quad . \quad . \quad . \quad . \quad (94)$$

The transformed band-pass frequency (which is a com-

binatíon of a low-pass and a high-pass normalized frequency) then becomes

$$\Omega = \frac{\omega_0}{\omega_B}\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right) \quad . \quad . \quad . \quad . \quad (95)$$

where

$$\omega_B = 2\pi\sqrt{(B_aB_p)}; \quad \omega_0 = 2\pi f_0 \quad . \quad . \quad (96)$$

It is important to note that, in order to use the Tables, eqn. (94) must be satisfied. In most practical cases it should not be too difficult to change slightly the band-edge frequencies so that their products are equal. The necessary change of specification should be made judiciously, preserving the most important requirements of the proposed filter.

The typical π- and T-sections of a band-pass filter are shown in Fig. 8, and the actual network components are given by the following equations.

$$\left.\begin{aligned} C_{(2s\mp1)\pi} &= \frac{1}{R^2}L_{(2s\mp1)T} = \frac{a_{2s\mp1}}{\omega_B R} \\[2mm] L_{(2s\mp1)\pi} &= R^2C_{(2s\mp1)T} = \frac{\omega_B R}{\omega_0^2 a_{2s\mp1}} \\[2mm] L_{(2s)\pi} &= R^2C_{(2s)T} = \frac{Ra_{2s}}{\omega_B} \\[2mm] C_{(2s)\pi} &= \frac{1}{R^2}L_{(2s)T} = \frac{\omega_B}{\omega_0^2 Ra_{2s}} \\[2mm] L'_{(2s)\pi} &= R^2C'_{(2s)T} = \frac{\omega_B Ra_{2s}}{\omega_0^2\Omega_{0s}^2} \\[2mm] C'_{(2s)\pi} &= \frac{1}{R^2}L'_{(2s)T} = \frac{\Omega_{0s}^2}{\omega_B Ra_{2s}} \end{aligned}\right\} \quad . \quad . \quad (97)$$
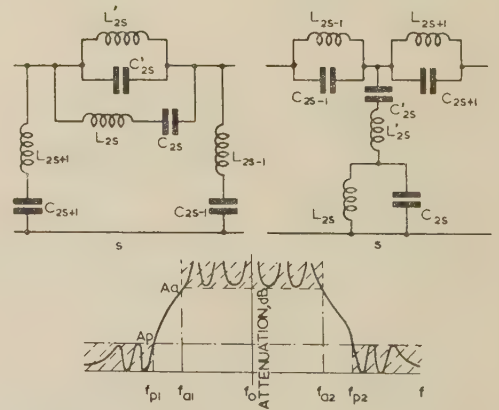
Fig. 9.—Typical band-stop filter sections (π- and T-configurations) and insertion-loss response.

(d) Band-Stop Filters (Fig. 9).

Power is to be passed with at most A_p decibels variation, within the ranges $0 \leqslant f \leqslant f_{p1}$ and $f_{p2} \leqslant f < \infty$.

Power is to be suppressed by at least A_a decibels, within the range $f_{a1} \leqslant f \leqslant f_{a2}$.

Here

$$\left.\begin{aligned} f_{p2} - f_{p1} &= B_p \\ f_{a2} - f_{a1} &= B_a \\ k &= \frac{B_a}{B_p} \end{aligned}\right\} \quad . \quad . \quad . \quad (98)$$

Again we require that

$$f_0^2 = f_{p1}f_{p2} = f_{a1}f_{a2}$$

as in eqn. (94).

The transformed band-stop frequency becomes now (again a combination of low-pass and high-pass corresponding terms):

$$\Omega = \cfrac{-1}{\dfrac{\omega_0}{\omega_B}\left(\dfrac{\omega}{\omega_0} - \dfrac{\omega_0}{\omega}\right)} \left.\vphantom{\begin{array}{c}a\\b\\c\end{array}}\right\} \quad . \quad . \quad . \quad (99)$$

$$\omega_B = 2\pi\sqrt{(B_p B_a)}; \quad \omega_0 = 2\pi f_0$$

The typical π- and T-sections of a band-stop filter are shown in Fig. 9, and the actual network components are given by the following equations.

$$\left.
\begin{aligned}
C_{(2s)T} &= \frac{1}{R^2}L_{(2s)\pi} = \frac{a_{2s}}{\omega_B R \Omega_{0s}^2} \\[2mm]
L_{(2s)T} &= R^2 C_{(2s)\pi} = \frac{\omega_B R \Omega_{0s}^2}{\omega_0^2 a_{2s}} \\[2mm]
C'_{(2s)T} &= \frac{1}{R_2}L'_{(2s)\pi} = \frac{\omega_B a_{2s}}{\omega_0^2 R} \\[2mm]
L'_{(2s)T} &= R^2 C'_{(2s)\pi} = \frac{R}{\omega_B a_{2s}} \\[2mm]
C_{(2s\mp1)T} &= \frac{1}{R^2}L_{(2s\mp1)\pi} = \frac{1}{\omega_B R a_{2s\mp1}} \\[2mm]
L_{(2s\mp1)T} &= R^2 C_{(2s\mp1)\pi} = \frac{\omega_B R a_{2s\mp1}}{\omega_0^2}
\end{aligned}
\right\} \quad . \quad . \quad (100)$$

The series arms of a T-configuration of band-pass or band-stop filters (see Figs. 8 and 9) may occasionally be rather difficult to realize since the ratio L_{2s}/C_{2s} may be inconvenient practically. In such cases it will be found necessary to use an equivalent network such as that shown in Fig. 10. The new series component values for the π-configuration, $L_{2s,1}$, $L_{2s,2}$ and $C_{2s,1}$, $C_{2s,2}$, can be found from the following formulae.

$$\text{Let} \quad r_s = \begin{cases} 1 + \dfrac{\omega_B^2}{2\omega_0^2\Omega_{0s}^2} \text{ for a band-pass filter} \\[3mm] 1 + \dfrac{\omega_B^2\Omega_{0s}^2}{2\omega_0^2} \text{ for a band-stop filter} \end{cases} \quad . \quad (101)$$

and

$$\left.
\begin{aligned}
\Omega_{\infty s,1}^2 &= r_s - \sqrt{(r_s^2 - 1)} \\
\Omega_{\infty s,2}^2 &= r_s + \sqrt{(r_s^2 - 1)}
\end{aligned}
\right\} \quad . \quad . \quad . \quad (102)$$

Then

$$\left.
\begin{aligned}
L_{2s,1} &= \frac{L'_{2s}}{1 + \Omega_{\infty s,1}^2} = L_{2s,2}\Omega_{\infty s,2}^2 \\[2mm]
C_{2s,1} &= C'_{2s}(1 + \Omega_{\infty s,2}^2) = \frac{1}{\Omega_{\infty s,1}^2}C_{2s,2}
\end{aligned}
\right\} \quad . \quad (103)$$

In similar ways a shunt branch of the T-configuration band-pass or band-stop filter can be transformed, if required, as shown in Fig. 10. Then

$$\left.
\begin{aligned}
L_{2s,1} &= L'_{2s}(1 + \Omega_{\infty s,2}^2) = \frac{1}{\Omega_{\infty s,1}^2}L_{2s,2} \\[2mm]
C_{2s,1} &= \frac{C'_{2s}}{1 + \Omega_{\infty s,1}^2} = \Omega_{\infty s,2}^2 C_{2s,2}
\end{aligned}
\right\} \quad . \quad (104)$$

It should be noted that the quantities $\Omega_{\infty s,1}^2$ and $\Omega_{\infty s,2}^2$ are
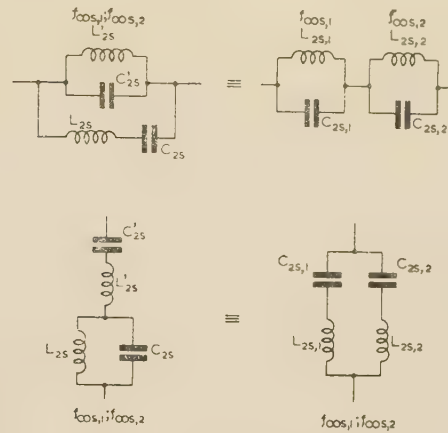
Fig. 10.—Equivalent networks.

related directly to the frequencies of infinite loss of a filter, as shown in Fig. 10. Thus:

$$\left.
\begin{aligned}
\Omega_{\infty s,1} &= \frac{f_{\infty s,1}}{f_0} \\[2mm]
\Omega_{\infty s,2} &= \frac{f_{\infty s,2}}{f_0}
\end{aligned}
\right\} \quad . \quad . \quad . \quad . \quad (105)$$

These quantities are therefore rather important in the actual filter alignment, since in practice the values of inductance and capacitance required in a tuned circuit are best adjusted by checking tuning frequencies. This point will be further discussed in Part 3.

(9) CORRELATING THE FILTER DESIGN SPECIFICATIONS

As explained in Section 5, the symmetrical filters which can be designed by the present method have their response specified in terms of the bandwidth ratio, k, and the levels of attenuation, A_a and A_p, in the suppression bands and the pass bands, respectively. From eqn. (46) these three quantities, if properly adjusted, specify uniquely the number of sections required to obtain the desired response. Thus, eqn. (46) can be written:

$$A_a = 12 \cdot 041 + (2n + 1)f(k) + g(A_p) \quad . \quad (106)$$

where A_a and A_p are in decibels. In Fig. 11, A_a is plotted for $n = 1$ and $n = 2$ as a function of the bandwidth ratio k, when $A_p = 1\,\text{dB}$. For any other value of the permissible distortion in the pass band, the whole curve as plotted will be shifted parallel to itself upwards (when $A_p > 1$) or downwards (when $A_p < 1$) by the amount specified by the function $g(A_p)$ in eqn. (106). The value of this shift is shown on the side of the graph and is designated 'A_p-scale'.

Thus, to correlate the specifications of the required filter, one has to decide first on the value of k corresponding to the type of filter, as explained in the preceding Section.

For a low-pass and band-pass filter:

$$k = \frac{B_p}{B_a} \qquad \text{(See Figs. 6 and 8)}$$

For a high-pass and band-stop filter:

$$k = \frac{B_a}{B_p} \qquad \text{(See Figs. 7 and 9)}$$

Next, the required value of the attenuation of the unwanted frequencies, i.e. A_a decibels, will determine a unique point in the
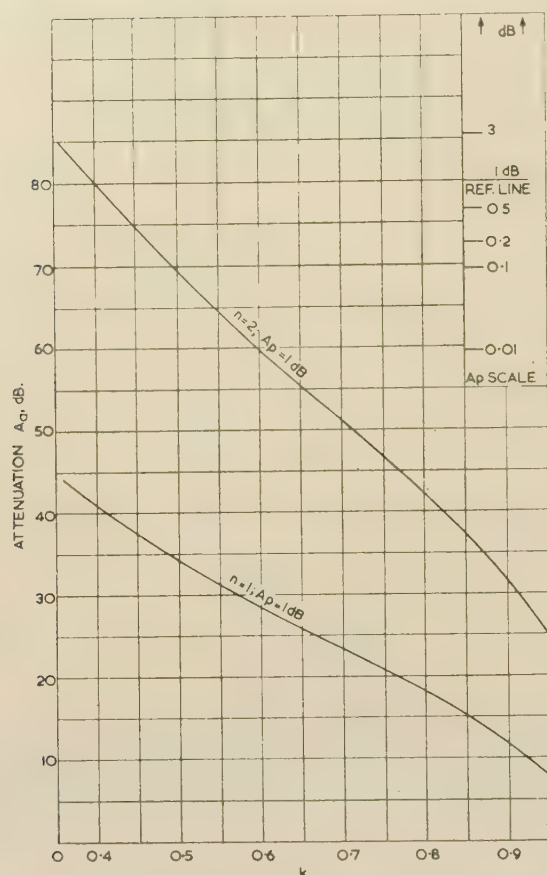
Fig. 11.—The correlation of the filter design specifications for Chebyshev behaviour in the pass and attenuation bands.

A_a, k plane in Fig. 11. Such a point will not, in general, lie on either of the $A_p = 1$ curves (i.e. for $n = 1$ or $n = 2$). Thus, supposing $k = 0.65$ and $A_a = 45$ dB, then for $n = 1$ the maximum distortion in the pass band is $A_p > 10$ dB, while for $n = 2$ it is $A_p = 0.1$ dB. Obviously, then, the filter which meets such specifications for k and A_a will have to consist of two sections.

It is advisable to reproduce the A_p scale on a separate piece of preferably stiff paper. Then (in the example above) place the scale along the ordinate $k = 0.65$ and slide it vertically until the reference line ($A_p = 1$ dB) coincides with the given n curve. For the curve $n = 2$, and for $A_a = 45$ dB, $A_p = 0.1$ dB; for $n = 1$, the same value of A_a is above the 3 dB line on the A_p-scale.

It will be seen in the next Section that the values of the normalized filter coefficients are given for a series of values of A_p (namely all the values shown on the 'A_p-scale' in Fig. 11 between 0.01 dB and 3 dB). It is therefore advisable to choose a value of A_p corresponding exactly to one of the marked values on the A_p-scale; this can almost always be arranged conveniently, e.g. by increasing slightly the value of A_a, or even, if conditions permit, the value of k. Otherwise, it will be necessary to interpolate between the curves described in the next Section, and this is better avoided.

(10) VALUES OF THE NORMALIZED SYMMETRICAL LADDER COEFFICIENTS FOR $n = 1$ AND $n = 2$

Once n (the number of sections), k (the bandwidth ratio) and A_p (maximum permissible deviation in the pass band) have been

chosen as explained in the preceding Section, the values of the ladder coefficients a_s of a symmetrical filter with equal terminations can be computed from the positions of the poles (i.e. $1/\Omega_{0s}$) and the zeros (i.e. α_s, ρ_s^2 and α_0) of the corresponding insertion-loss function in the complex frequency plane. The particular choice of the insertion-loss function and the resulting co-ordinates of the poles and zeros have been explained and determined in Section 5. Thus, from eqns. (52) the values of α_s and ρ_s^2 are given in terms of α_0, Ω_{0s}^2 and k, while α_0 is, from

Table 1

CHARACTERISTIC CONSTANTS GOVERNING THE SHAPE OF THE INSERTION-LOSS FUNCTION

k	$n = 1$	$n = 2$	
	Ω_{01}	Ω_{01}	Ω_{02}
0.35	0.51646	0.35515	0.56434
0.40	0.55352	0.38230	0.60388
0.45	0.58894	0.40880	0.64129
0.50	0.62297	0.43503	0.67683
0.55	0.65606	0.46134	0.71103
0.60	0.68845	0.48811	0.74388
0.65	0.72048	0.51572	0.77585
0.70	0.75243	0.54476	0.80712
0.75	0.78460	0.57590	0.83760
0.80	0.81752	0.61026	0.86784
0.85	0.85185	0.64968	0.89809
0.90	0.88874	0.69791	0.92877
0.95	0.93106	0.76532	0.96084

eqns. (50) and (51), dependent on k, n and A_p. Table 1 gives the values of Ω_{0s} for $n = 1$ and $n = 2$ and for

$$k = 0.35(0.05)0.95$$

Interpolation for intermediate values of A_p should be avoided, since these values were not chosen at equal intervals; their choice was governed by practical utility and for completeness. It will be noted that for higher values of k and for the lower values of A_p, the values of a_s are not given fully in Figs. 12–18; for these values of the specified parameters A_p, k and n the resultant filters cannot be realized by a simple ladder structure without mutual reactances, since then some of the values of the ladder coefficients may become negative. On the other hand, it will be seen in Fig. 11 that, especially for $n = 1$, higher values of k and low values of A_p are precluded.

It was shown in Section 7 that, given the position of the poles and zeros of a symmetrical equally terminated network, the values of its ladder coefficients can be explicitly derived. In particular, eqns. (75), (76), (82) and (83) give values of a_s for $n = 2$, and eqn. (84) gives values of a_s for $n = 1$ (note that in this last case $a_1 = a_3$). All these coefficients are plotted for various values of A_p as functions of the bandwidth ratio k. Thus, Figs. 12 and 13 give $a_1 = a_3$ and a_2, respectively, for $n = 1$, while Figs. 14, 15, 16, 17 and 18 give a_1, a_2, a_3, a_4 and a_5, respectively, for $n = 2$. In some cases, the curves are plotted in two separate parts. Thus, for $n = 2$, a_2 for $0.01 \leqslant A_p \leqslant 0.5$ (Fig. 15) is plotted in the upper part of the Figure (scale on the right), while a_2 for all A_p is plotted below (scale on the left) for clarity, since, while in the first place the curves move upwards when A_p increases, they subsequently move downwards and overlap. The scales are chosen so that the values of a_s can be read with 3-figure (i.e. slide-rule) accuracy, which should be sufficient for most practical purposes. Similarly, in Fig. 17, curves for individual values of a_s crowd too much together and even cross each other. Such troublesome parts are enlarged,
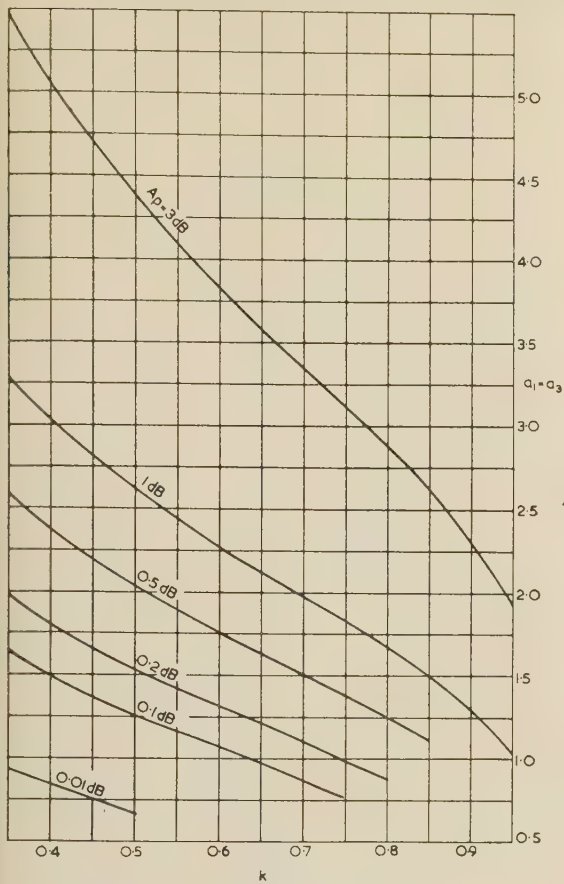
Fig. 12.—Ladder coefficients $a_1 = a_3$ $(n = 1)$.
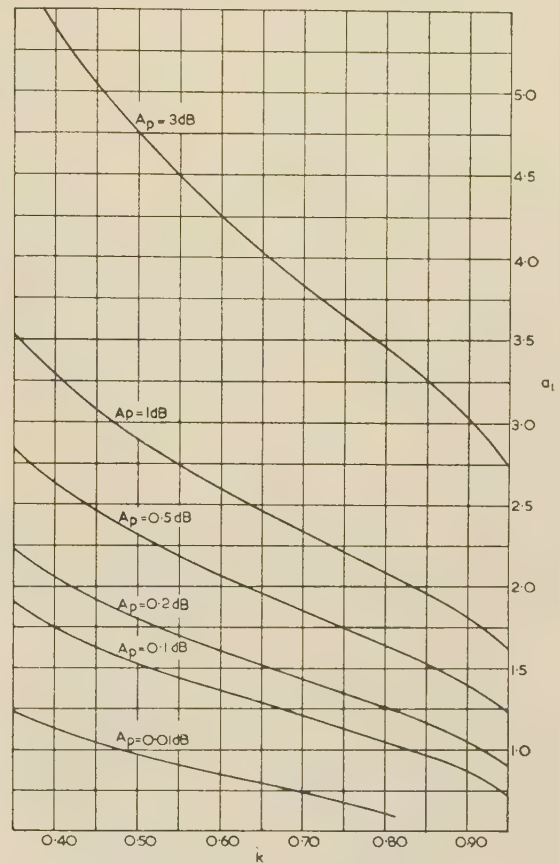
Fig. 14.—Ladder coefficient a_1 $(n = 2)$.
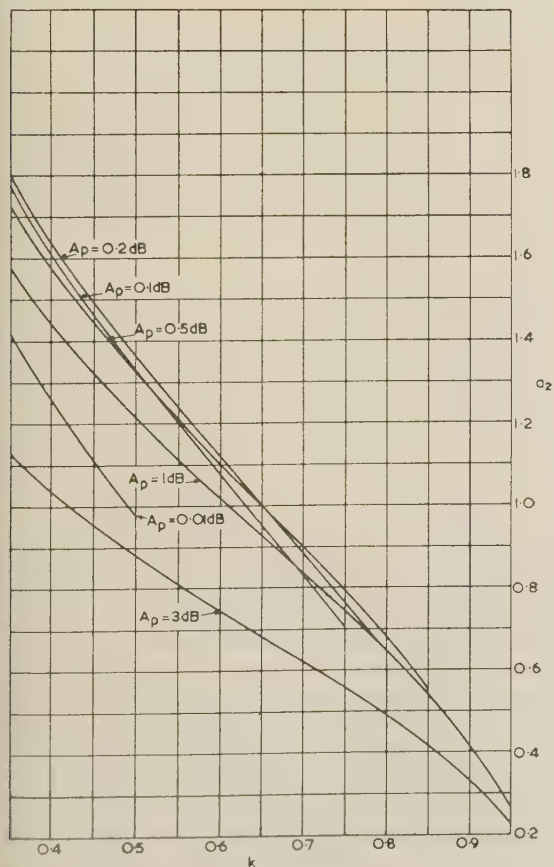
Fig. 13.—Ladder coefficient a_2 $(n = 1)$.
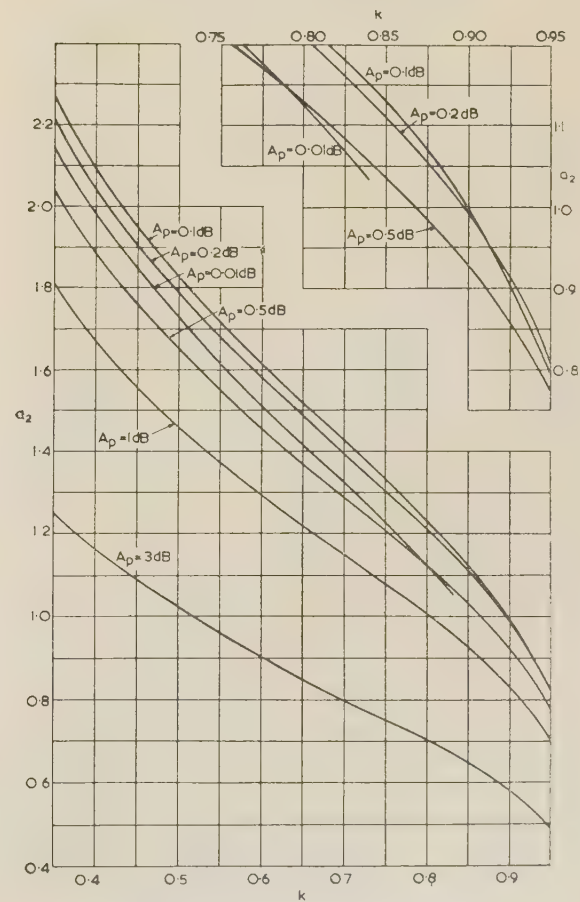
Fig. 15.—Ladder coefficient a_2 $(n = 2)$.
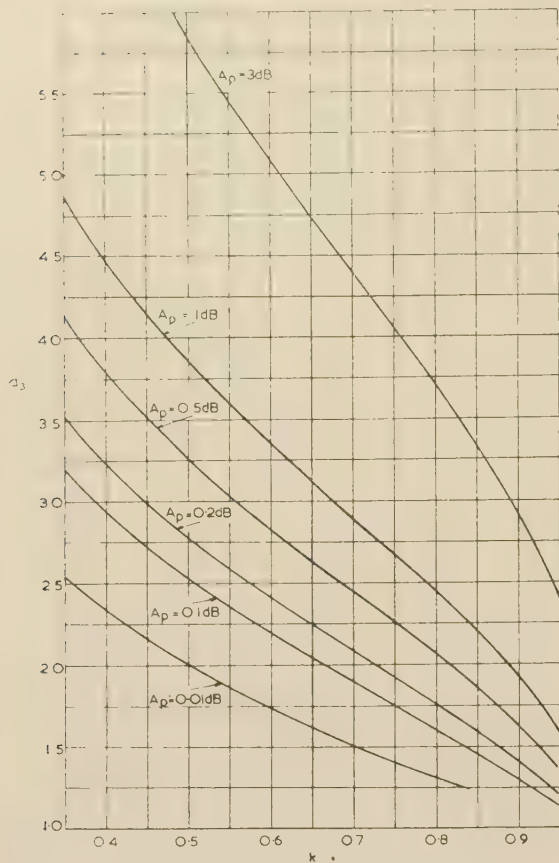
Fig. 16.—Ladder coefficient a_3 ($n = 2$).

Fig. 17.—Ladder coefficient a_4 ($n = 2$).

Fig. 18.—Ladder coefficient a_5 ($n = 2$).

and the separation or magnification of curves is clearly indicated; care should be taken in using the proper vertical scales when reading off the particular values of a_s.

As explained in Section 8, the ladder coefficients a_s and the frequencies of infinite loss Ω_{0s} give directly the actual network components (i.e. the inductances and capacitances) depending on the type of filter, its characteristic frequencies (e.g. bandwidth, carrier) and the value of the generator load resistance.

Part 3.   PRACTICAL CONSIDERATIONS

This Part of the paper is divided into two Sections; in the first we consider certain practical points which arise when the design data are used to obtain the actual filter. The second Section presents a method of estimating reasonable Q-factors.

An actual example of design is given in the Appendix.

(11) PRACTICAL FILTER ALIGNMENT

The design data presented in Part 2 in graphical form were computed for purely reactive filters. Unavoidable ohmic losses in the components will result in a certain distortion of the prescribed filter response. There exist methods of accounting for losses either by assigning them equally to all components or by using more elaborate loss relations; both cases are fully dealt with by Darlington.[6] It seems, then, that in most cases to compensate for losses in a filter designed for equal terminations the ladder coefficients $a_{2s\mp1}$ should be decreased, while the coefficients a_{2s} should be increased. The amount of this increase or decrease varies greatly, even when assigning equal losses to all components (e.g. for the so-called 'predistorted filter'; see Belevitch[2]).

On the other hand, especially in the case of band-pass and

nd-stop filters, the components must finally be aligned, even
nen computed with losses. Some points will now be discussed
ich should be remembered during such an alignment.

The alignment of a low-pass or high-pass filter is quite straight-
ward because of the relatively simple structure of these net-
rks; usually, one needs only to adjust properly the components
rest to the input and output terminals after the resonant
cuits have been properly tuned.

However, the alignment becomes much more involved for
nd-pass and band-stop filters, since there we have not only
tune resonant circuits to proper frequencies, but also to
ure that the L/C ratios are correct, thereby ensuring proper
ing to the mid-band frequency and preventing the pass-band
iation from exceeding the prescribed limit A_p decibels. The
rent component losses (e.g. finite Q-factors of the coils) in
t act to the benefit of the designer in the pass band by
oothing out the insertion-loss variations there, but they also
er effective attenuation in the suppression bands. It is there-
advisable to choose, in practice, slightly lower pass-band
iation (i.e. A_p), a slightly higher value of k and, if possible,
her attenuation A_a in the suppression bands than those
ually required. The presence of unavoidable losses will then
ng the resultant response within prescribed limits, after
ieful alignment of components.

The most important but also most difficult to align are the
es branches in the π-configuration (Fig. 8) and the shunt
nches in the T-configuration (Fig. 8), i.e. those branches of
band-pass or band-stop filter which involve the coefficients
. These should first be aligned separately so as to produce
ximum rejection at their proper frequencies $f_{\infty s, 1, 2}$ and mini-
m rejection at the mid-band frequency f_0. If an equivalent
cuit is used (see Section 8 and Fig. 10) in the form of two anti-
onant circuits in series, the alignment is best performed by
nnecting these two circuits in parallel and tuning such a
mbination to maximum rejection at $f_{\infty s, 1, 2}$ and minimum
ection at f_0, thus ensuring proper values of inductance and
pacitance. When using the equivalent circuit for the series
anch of a π-configuration (see Fig. 10), care should also be
en to minimize the stray capacitance between the earth lead
d the junction point between the two anti-resonant circuits.
e other stray capacitances are easily accounted for by adjusting
values of the $C_{2s \mp 1}$ condensers.

The alignment procedure is greatly simplified if a cathode-ray-
be display unit is available, capable of producing, simul-
eously, the filter response over the whole frequency range
uired. However, if the point-by-point alignment method is
ed, it is a good policy to compute the positions of the minima
the pass band (i.e. frequencies where the loss is minimum),
ich for a low-pass filter correspond to the frequencies

$$f_{0s} = \Omega_{0s} f_B$$

d for a band-pass filter to

$$\left.\begin{array}{c} f_{0s,1} \\ f_{0s,2} \end{array}\right\} = \sqrt{[f_0^2 + (\tfrac{1}{2} f_B \Omega_{0s})^2]} \mp \tfrac{1}{2} f_B \Omega_{0s}$$

ese frequencies, as well as those corresponding to infinite
ection, should be strictly adhered to. In the case of a high-
ss or a band-stop filter the frequencies of minimum loss are
tained from the above formulae by replacing the Ω_{0s} by their
procals.

) A REASONABLE ESTIMATE OF MINIMUM ALLOWED
Q-FACTORS

As stressed in the preceding Section, the method of filter design
scribed here makes no allowance for the ohmic losses of com-
nents. Several methods are, in fact, available to adjust the

component values accordingly; these 'predistortion' methods
are described fully by Darlington.[6] A symmetrical filter, how-
ever, when designed with the predistortion techniques, ceases
to be symmetrical in the proper sense of the definition (see
Section 3).

The easiest predistortion technique consists of shifting the
real frequency axis in the complex frequency p-plane by a small
amount which eventually is made to be equal to the reciprocal
of the Q-factor of the coils. The shift can in no circumstances
be greater than α_n [which is equal to the real part of the zero
nearest to the real frequency axis in Fig. 4; see also eqn. (33)]. The
usual practice is to take this shift to be equal to $\alpha_n/3$.

It was found experimentally, on the basis of long practice in
the laboratories, that by taking

$$Q_{LP} \geqslant \frac{10}{\alpha_n} \quad . \quad . \quad . \quad . \quad . \quad (107)$$

for a low-pass filter, the resultant adjustment of components can
be easily performed even when these components have been
obtained without the use of the predistortion techniques. The
Q-factors required are thus about three times as large as those
obtained when predistortion is used.
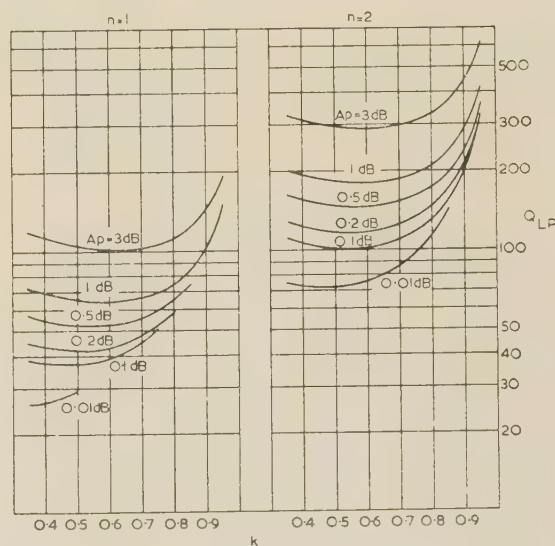
Fig. 19.—Data for the estimation of minimum Q-factors.

In Fig. 19 the minimum Q-factors are plotted from con-
dition (107) as functions of the number of sections, n, and the
bandwidth ratio, k, for the values of A_p used in the component
graphs (Figs. 12–18).

It should be stressed that these values of Q-factors are to be
regarded as a guide only.

For band-pass or band-stop filters the Q-factors become

$$Q_{BP} \text{ or } Q_{BS} = Q_{LP} \frac{f_0}{f_B} \quad . \quad . \quad . \quad (108)$$

where f_0 = centre frequency
$f_B = \sqrt{(B_a B_p)}$

as in eqns. (96) and (99).

(13) ACKNOWLEDGMENTS

(14) BIBLIOGRAPHY

(1) BADER, W.: 'Polynomvierpole vorgeschriebener Frequenz-abhängigkeit', *Archiv für Elektrotechnik*, 1940, **34**, p. 181.

(2) BELEVITCH, V.: 'Tchebyshev Filters and Amplifier Networks', *Wireless Engineer*, 1952, **29**, p. 106.
'Topics in the Design of Insertion Loss Filters', *Transactions of the Institute of Radio Engineers*, 1955, CT2, No. 4.

(3) BENNETT, W. R.: U.S. Patent No. 1849656 (March, 1952).

(4) CAUER, W.: 'Theorie der Linearen Wechselstromschal-tungen', 2 Auflage (Akademie-Verlag, Berlin, 1954).

(5) COCCI, G.: 'Funzioni di trasmissione di quadripoli di pure reattanze inseriti tra due resistenze', *Alta Frequenza*, 1938, **7**, p. 804.
'Progetto di quadripoli di pure reattanze con funzione di trasmissione data', *ibid.*, 1941, **10**, p. 470.
'Filtri con numero minimo di elementi', *ibid.*, 1942, **11**, p. 482.

(6) DARLINGTON, S.: 'Synthesis of Reactance Four-Poles which produce Prescribed Insertion Loss Characteristics, including Special Applications to Filter Design', *Journal of Mathematics and Physics*, 1939, **18**, p. 257.

(7) DISHAL, M.: 'The Design of Dissipative Band-Pass Filters producing Desired Exact Amplitude Frequency Characteristics', *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 1050.

(8) GREEN, E.: 'Amplitude Frequency Characteristics for Ladder Networks', Marconi's Wireless Telegraph Co. Ltd.

(9) JAHNKE, E., and EMDE, F.: 'Tables of Functions with Formulae and Curves' (Dover Publications, New York, 1954).

(10) MILNE-THOMSON, L. M.: 'Die elliptischen Funktionen von Jacobi' (Springer Verlag, Berlin, 1931).

(11) NORTON, E. L.: 'Constant Resistance Networks with Applications to Filter Groups', *Bell System Technical Journal*, 1937, **16**, p. 178.

(12) ORCHARD, H. J.: 'Formulae for Ladder Filters', *Wireless Engineer*, 1953, **30**, p. 3.

(13) PILOTY, H.: 'Kanonische Kettenschaltungen für Reaktanz-vierpole mit vorgeschriebenen Betriebseigenschaften', *Telegraphen-Fernsprech-Technik*, 1940, **29**, p. 249.

(15) APPENDIX: AN EXAMPLE ON THE USE OF THE TABLES

A required low-pass filter is to be equally terminated in 1 kilohm, and should pass frequencies up to $2 \cdot 8$ Mc/s with an attenuation not greater than $\pm 0 \cdot 05$ dB; it should suppress all signals above 4 Mc/s by at least 40 dB.

The filter specification can thus be written (see Fig. 1):

$$f_p = B_p = 2 \cdot 8 \text{ Mc/s}, \quad A_p = 0 \cdot 1 \text{ dB}$$
$$f_a = B_a = 4 \text{ Mc/s}, \quad A_a \geqslant 40 \text{ dB}$$

so that
$$k = \frac{B_p}{B_a} = 0 \cdot 7$$

From Fig. 11 we find that for $k = 0 \cdot 7$ and $A_p = 0 \cdot 1$, A just slightly above 40 dB for $n = 2$. The complete filter spec cation is therefore

$$n = 2 \text{ sections (5 branches)}$$
$$k = 0 \cdot 7; \quad A_p = 0 \cdot 1 \text{ dB}; \quad A_a > 40 \text{ dB}$$

The characteristic constants Ω_{0s} governing the shape of insertion-loss function are found in Table 1:

$$\Omega_{01} = 0 \cdot 54476; \quad \Omega_{02} = 0 \cdot 80712$$

while the ladder coefficients for the designated values of k and are found from Figs. 14, 15, 16, 17 and 18:

$$a_1 = 1 \cdot 21 \qquad a_2 = 1 \cdot 425$$
$$a_3 = 1 \cdot 90 \qquad a_4 = 1 \cdot 05$$
$$a_5 = 0 \cdot 92$$

To obtain the actual low-pass filter components it is necessa to evaluate the characteristic filter frequency:

$$f_B = \sqrt{(B_p B_a)} = 3 \cdot 3466 \text{ Mc/s}$$

We choose the configuration in order to account for the str capacitances across the filter terminals. The component valu can be computed with slide-rule accuracy from the formulae (8 given in Section 8. For $R = 10^3$ ohms,

$$\omega_B \simeq 21 \cdot 0 \times 10^6$$
$$\Omega_{01}^2 \simeq 0 \cdot 297 \qquad \qquad \Omega_{02}^2 \simeq 0 \cdot 651$$

and from a_s as found from the graphs,

$$C_1 = 58 \text{ pF} \qquad L_2 = 68 \, \mu\text{H} \qquad C_2 = 10 \text{ pF}$$
$$C_3 = 90 \text{ pF} \qquad L_4 = 50 \, \mu\text{H} \qquad C_4 = 29 \text{ pF}$$
$$C_5 = 44 \text{ pF}$$

In Fig. 19 we see that the Q-factor of the coils should be great than 110.

The filter will usually need slight adjustments, but the actu alignment will be simplified if the following points are careful watched:

(a) The shunt capacitors C_1 and C_5 need to be adjusted to accoun for the stray, the input and the output network capacitances.
(b) The series-resonant arms need careful tuning to prop frequencies of rejection with the proper L_{2s}/C_{2s} ratios retained.

The final alignment should aim at producing the response shap postulated, so that the frequencies of 'zero' and 'infinite' lo should be carefully adjusted. These are

$$f_{\infty s} = \frac{f_B}{\Omega_{0s}} \qquad \qquad f_{0s} = f_B \Omega_{0s}$$

Thus, in megacycles per second,

$$f_{\infty 1} = 6 \cdot 1433 \qquad f_{01} = 1 \cdot 8231$$
$$f_{\infty 2} = 4 \cdot 1464 \qquad f_{02} = 2 \cdot 7011$$

MODE SEPARATION AT THE π-MODE IN A DIELECTRIC LOADED WAVEGUIDE CAVITY

By G. B. WALKER, Ph.D., and N. D. WEST, B.Sc.(Eng.).

SUMMARY

The dispersion characteristics of periodically loaded waveguides are examined. As is well known, a loaded waveguide has the properties of a filter having infinitely many stop bands, which occur when the phase change per section for the wave is a multiple of π. For linear accelerator and other applications it is sometimes desirable to operate the structure in the π-mode but the resulting zero group velocity raises considerable difficulties.

In the first part of the paper it is shown how a finite group velocity the π-mode frequency can be obtained in a waveguide loaded with solid dielectric discs. This is accomplished by making the discs reflectionless at the required frequency; the stop band being thus eliminated, the structure can be regarded as a confluent band-pass filter.

To permit the passage of an electron beam it is necessary for the discs to have central holes. Evanescent and other propagating modes are set up, but it is shown in the latter part of the paper that it is still possible to operate the structure in the π-mode with a finite group velocity. This result cannot be obtained if the guide is loaded by thin metal discs, as used in most present-day linear accelerators.

LIST OF PRINCIPAL SYMBOLS

a = Circular-waveguide radius.
L = Distance between loading obstacles.
$2L_a$ = Length of air-filled region of waveguide.
$2L_d$ = Thickness of dielectric disc.
β = Phase-change coefficient.
β_a = Phase-change coefficient in air-filled guide.
β_d = Phase-change coefficient in dielectric-filled guide.
β_k = Radial phase-change coefficient in circular waveguide.
ϕ = Phase change per period in periodically loaded guide.
Z_a = Characteristic wave impedance in air-filled guide.
Z_d = Characteristic wave impedance in dielectric-filled guide.
μ_0, ϵ_0 = Permeability and permittivity, respectively, of free space.
ϵ = Dielectric constant of the dielectric disc.
ω = Angular frequency.
f_c = Cut-off frequency in empty guide.
f_m = Frequency at which the characteristic wave impedances in a dielectric-filled and an empty waveguide are equal.
c = Velocity of light.
v_g = Group velocity.

(1) INTRODUCTION

Waveguide structures capable of supporting a wave with phase velocity less than, or equal to, the velocity of light, have received considerable attention in recent years. For linear-accelerator applications the slowing of the wave, so that it may interact with

high-energy electrons, is usually obtained by loading a hollow circular cylindrical waveguide with metal irises at regular intervals. A similar result can be obtained by loading with dielectric discs as in the case of the travelling-wave linear accelerator of Harvey, Mullet and Walkinshaw.[1,2,3]

There are certain advantages to be gained by the use of dielectric-disc loading. With dielectric materials such as titania it is possible to have lower losses, and thus a higher accelerating efficiency, than can be obtained in all-metal structures,[4] but an equally important feature is that structures with better dispersion characteristics can be produced.

In connection with the design of a resonant accelerator an investigation has been made into the mode spectrum of dielectric-disc loaded cavities operated in an E-mode, and it is the purpose of the paper to indicate how it is possible to achieve a higher mode separation in the region of π-mode operation than can be had by the use of metal irises.

(2) GENERAL THEORY

(2.1) Filter Characteristic of Periodic Structures

As shown by Slater[5] a periodically loaded waveguide behaves as a band-pass filter, having, in general, infinitely many pass and stop bands associated with each mode type. It will propagate a wave disturbance which can be analysed into a series of space harmonics having phase-change coefficients given by $\beta_n = \beta_0 \pm 2\pi n/L$ where L is the periodic length (distance between loading obstacles), n is an integer and β_0 is taken to be the phase-change coefficient of the fundamental harmonic. As the composite wave consists of harmonics having phase changes per period differing from each other by $2\pi n$, it does not itself have a unique phase change per period. It is, however, convenient to attribute a phase change of ϕ to the composite wave equal to that of a reference harmonic, which is chosen in such a way that ϕ lies between 0 and π, in the first pass band, π and 2π in the second pass band, and so on. In many periodic structures this reference harmonic has the greatest amplitude. A plot of frequency against ϕ therefore appears as in Fig. 1, where a discontinuity at a phase change of π indicates the first stop band, with other stop bands occurring when ϕ is a multiple of π.

An important quantity is the group velocity v_g, given by $d\omega/d\beta$, which is the same for all the space harmonics at any one frequency, and is proportional to the slope of the tangent to the curve. The phase velocity ω/β is, of course, proportional to the slope of the radius vector from the origin to the operating point on the curve.

(2.2) Condition for Resonance

If reflecting end walls are placed at points midway between loading obstacles so as to form a waveguide cavity containing n sections, a resonance is possible when the total phase change in a progressive wave is an integral multiple of π. If m is an integer then for resonance, $n\phi = m\pi$. Thus, in general, there are $n + 1$ values of ϕ which give resonances between 0 and π, π and

Fig. 1.—Typical dispersion curve for periodically loaded waveguide.

The abscissa is the phase change per section and the curve is discontinuous at the value π. Ten equal intervals have been marked off between 0 and π to illustrate resonant modes of a cavity containing ten sections. Only a small frequency difference occurs between the π-mode and its nearest neighbour.

2π, etc. The $n + 1$ resonant modes in the first pass band are shown in Fig. 1 for a 10-section cavity. It can be seen that the frequency separation between adjacent modes is a minimum at the ends of the pass bands where the group velocity falls to zero. The mode separation will obviously also decrease as the number of sections increases.

(2.3) Choice of Operating Mode

Several workers[6] have noticed that there is an advantage to be gained from operating a resonant linear accelerator in the π-mode. The advantage arises from the fact that the amplitude of the wave which couples or resonates with the particle is suddenly increased in amplitude when the π-mode is reached. This can be shown by a consideration of the various space harmonics which exist in the resonant accelerator (see Appendix 8).

It is possible to make use of this mode of operation if the accelerator is relatively short, so that the adjacent modes are sufficiently far away to present no difficulty in the feeding of power to the accelerator. To make the best use of the available power sources, the cavity should be as long as possible,[7] and it is found, in practice, that the limit to the possible length in π-mode operation in present accelerators is set by the mode separation.[8]

For this reason some workers[9] have preferred to operate a resonant accelerator in the region of the $\pi/2$ mode ($\pi/2$ rad phase change per section), where much longer cavities can be used before interference between adjacent modes becomes a difficulty. It should be noticed that $\pi/2$-mode operation involves the use of twice as many loading obstacles as the π-mode for a given cavity length, and consequently some of the advantage in mode separation to be gained by working in this region of large group velocity is lost.

In addition to the accelerating efficiency of the π-mode it has other, though less important, advantages. Fewer loading obstacles are required than in other modes of operation, thus leading to a simplification in manufacture. The large spacing between obstacles tends to reduce the multipactor effect sometimes found in accelerators,[10] by making half-cycle transit-time multipactoring impossible for fields of the order of magnitude found in accelerators.

(2.4) Condition for Disappearance of the Stop Band

Operation of the loaded resonator in the region of the π-mode has been shown to involve a small mode separation, arising from the zero group velocity at the π-mode. When the phase cha per section is π radians the reflections of a progressive wave fr successive loading obstacles add together in phase. In an infin loss-free structure the reflected wave builds up to become eq in magnitude to the incident wave, so that a pure standing w is set up having zero group velocity. If the loading obstacles reflectionless at the π-mode frequency there can be no build of a backward wave, and so it should be possible for a travell wave of finite group velocity to be propagated. Under suc condition there is no stop band at the point $\phi = \pi$, and the gu behaves like the confluent band-pass filter of electrical circ theory.[11] In the following Section it will be shown that suc structure can be produced by using solid dielectric disc loadi

(3) THE DIELECTRIC-DISC-LOADED WAVEGUIDE

(3.1) Derivation of Filter Characteristics

The transmission characteristic of an infinite length of wa guide loaded at regular intervals by solid dielectric discs (Fig
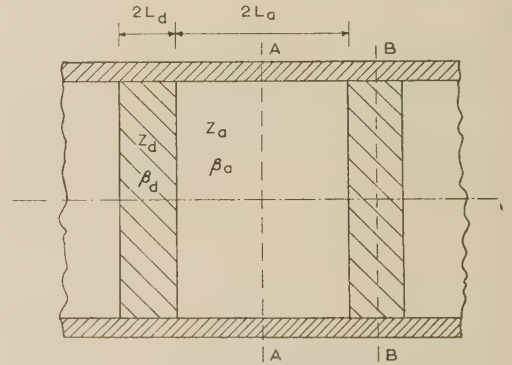
Fig. 2.—Waveguide loaded with dielectric discs.

can be deduced from a knowledge of the characteristic wa impedances and propagation coefficients of the air and dielect regions. The structure will be assumed to be lossless and propagate only one mode type, for which the wave impedan and phase-change coefficients of the air and dielectric regio are Z_a, Z_d, and β_a, β_d, respectively. The analysis proceeds as a cascade of identical networks.[11] It is convenient to divide guide into half-sections by reference planes placed midw between discs and through the centre of each disc. The syst of half-sections can then be regarded as a cascade of identi dissymmetrical networks connected on an image basis.

For the half-section between AA and BB in Fig. 2, let Z_{sc} a Z_{oc} be the input impedances at AA when a short-circuit or open-circuit is placed at BB. It can be shown that, for an infin line,

$$j \tan \phi/2 = \frac{Z_{sc}}{Z_{oc}}$$

where ϕ is the phase change per section as before.

Transforming impedances from BB to AA, Z_{oc} and Z_{sc} can found, and are given by

$$Z_{oc} = \frac{Z_a Z_d - Z_a^2 \tan \beta_a L_a \tan \beta_d L_d}{jZ_a \tan \beta_d L_d + jZ_d \tan \beta_a L_a}$$

$$Z_{sc} = \frac{Z_a Z_d + Z_a^2 \tan \beta_a L_a \cot \beta_d L_d}{-jZ_a \cot \beta_d L_d + jZ_d \tan \beta_a L_a}$$

om which,

an $\phi/2 =$

$$\left\{ \frac{[\tan \beta_d L_d + (Z_d/Z_a) \tan \beta_a L_a)][(Z_d/Z_a) \tan \beta_d L_d + \tan \beta_a L_a]}{[(Z_d/Z_a) \tan \beta_a L_a \tan \beta_d L_d - 1](Z_d/Z_a - \tan \beta_a L_a \tan \beta_d L_d)} \right\}^{1/2}$$

. . . . (1)

The limits of the stop bands occurring when ϕ is an odd ultiple of π are given by $\tan \phi/2 = \infty$, and are therefore

$$\tan \beta_a L_a \tan \beta_d L_d = Z_d/Z_a \quad . \quad . \quad . \quad . \quad (2)$$

$$\tan \beta_a L_a \tan \beta_d L_d = Z_a/Z_d \quad . \quad . \quad . \quad . \quad (3)$$

hile the limits when L is an even integral of π are given by n $\phi/2 = 0$ and are as follows:

$$\tan \beta_a L_a \cot \beta_d L_d = -Z_d/Z_a . \quad . \quad . \quad . \quad (4)$$

$$\tan \beta_a L_a \cot \beta_d L_d = -Z_a/Z_d . \quad . \quad . \quad (5)$$

The modes of particular interest are the π-modes, which are iven by the lowest-frequency solutions of eqns. (2) and (3). It is easy to see that for a disc-loaded waveguide the loading stacles are reflectionless if the characteristic wave impedance the air region is equal to that in the dielectric region, i.e. $= Z_d$. It is shown in the next Section that this condition n be obtained for an E-mode wave by the correct choice of uide dimensions (i.e. cut-off frequency fc), operating frequency d dielectric constant of the discs. The discs will only be flectionless at one frequency, but if the frequency chosen is at of one of the π-modes then the condition to produce con- ience of the $0–\pi$ and $\pi–2\pi$ pass bands is satisfied. If the condition $Z_a = Z_d$ is substituted in eqns. (2) and (3) it readily seen that the two π-modes have the same frequency, nich is another way of stating that the pass bands are confluent.

2) Matching of Characteristic Wave Impedance between Air- Filled and Dielectric-Filled Regions

For a uniform waveguide propagating a single mode type, it is ell known that the characteristic wave impedances are given by

$$Z_0 = \frac{1}{\epsilon}\sqrt{(\mu_0/\epsilon_0)}\sqrt{[\epsilon - (f_c/f)^2]} \text{ for an E-mode,}$$

and $$Z_0 = \sqrt{(\mu_0/\epsilon_0)}/\sqrt{[\epsilon - (f_c/f)^2]} \text{ for an H-mode}$$

It is obvious that for an H-mode in a given waveguide Z_0 must ways be greater when ϵ is unity than when ϵ is greater than unity. us for an H-mode wave it is not possible to achieve an impe- nce match between air-filled and dielectric-filled regions of a aveguide. However, for an E-mode this is not the case, and at frequency given by

$$f_m = f_c\sqrt{\frac{1 + \epsilon}{\epsilon}}$$

e characteristic wave impedances in the two regions are equal. nis condition was first noticed by Lamont.[12,13] In general, nere the regions are filled with dielectric of dielectric constants and ϵ_2, the matching frequency is given by

$$f_m = f_c\sqrt{\frac{\epsilon_1 + \epsilon_2}{\epsilon_1 \epsilon_2}}$$

3.2) Mode Separation in Dielectric-Disc-Loaded Waveguide Cavity

It is convenient first to designate the two π-modes, π_1 and π_2, nose frequencies are given by eqns. (2) and (3), respectively, nere π_1 has a node at AA in Fig. 2 and π_2 has a node at BB.

Which one of the π-modes is excited depends, therefore, on where the cavity is terminated, and for a given termination only one of the π-modes can be excited. The other modes in an n-section cavity are numbered according to the number of π-phase changes there are along the cavity length—those adjacent to the π-modes being the $(n + 1)$ and $(n - 1)$ modes.

It is difficult to derive an expression for frequency separation between the π-mode and its neighbour, because, in general, the zero slope of the dispersion curve at the π-mode makes a first- order approximation impossible. The frequency of the $(n + 1)$ and $(n - 1)$-modes can be determined by calculating the dis- persion curve in the region of the π-mode. When the pass bands become confluent a first-order approximation is possible for small frequency differences, because of the finite group velocity. It can be shown that the group velocity at the matched π-mode is given by

$$v_g = \frac{(\tan \beta_a L_a + \tan \beta_d L_d)\pi L f}{\left[\left(\frac{\omega}{c}\right)^2 (\tan \beta_a L_a + \tan \beta_d L_d)^2 (L_a + L_d)^2 (\epsilon + 1) - \left(\epsilon - \frac{1}{\epsilon}\right)^2\right]^{1/2}}$$

and that the mode separation Δf in terms of v_g is

$$\Delta f = \pm \frac{v_g}{2nL}$$

In linear accelerators it is customary to use a circular cylindrical loaded waveguide operated in the E_{01}-mode, and so the numerical work in this Section deals with this case.

For a circular waveguide, in eqn. (1),

$$\beta_a = \sqrt{(\omega^2/c^2 - \beta_k^2)}$$
$$\beta_d = \sqrt{(\omega^2/c^2\epsilon - \beta_k^2)}$$

and $$\frac{Z_d}{Z_a} = \frac{\beta_d}{\beta_a\epsilon}$$

where for an E_{01}-mode, if a is the guide radius, $\beta_k a$ is the first root of the zero-order Bessel function.

A comparison between dispersion curves for disc-loaded wave- guides which are matched and unmatched at the π-mode is shown in Fig. 3. In both waveguides the same π_1-mode frequency, disc spacing and dielectric constant are taken, but for the non- matching case the waveguide diameter has been adjusted so that the π-mode frequency is 3.9% higher than that required for matching. The group velocity at the π-mode for the confluent curve is seen to be the maximum value over the whole pass band.

Instead of presenting a series of dispersion curves in order to show how the mode separation varies as any one of the guide dimensions is varied, it is convenient to show the mode separation directly plotted against the relevant variable. Of the variables available, f, ϵ, a, L_a, and L_d, some are fixed for any particular accelerator design. For example, using an available low-loss dielectric material for the discs it is necessary to assume ϵ to be fixed. Similarly the operating frequency is generally determined by available high-power oscillators which, in turn, for a length of accelerator of constant phase velocity, determines the disc spacing. The remaining variable to satisfy the matching condi- tion is the tube radius a, which, when chosen, determines the disc thickness.

In Fig. 4 the frequency of the π_1-mode, which is the most efficient of the π-modes for a linear accelerator, is taken as fixed, and the variations of the π_2, $(n - 1)$ and $(n + 1)$ mode frequencies are plotted against the guide radius for a particular 10-section cavity.

If the cavity is terminated so that the operating mode is the
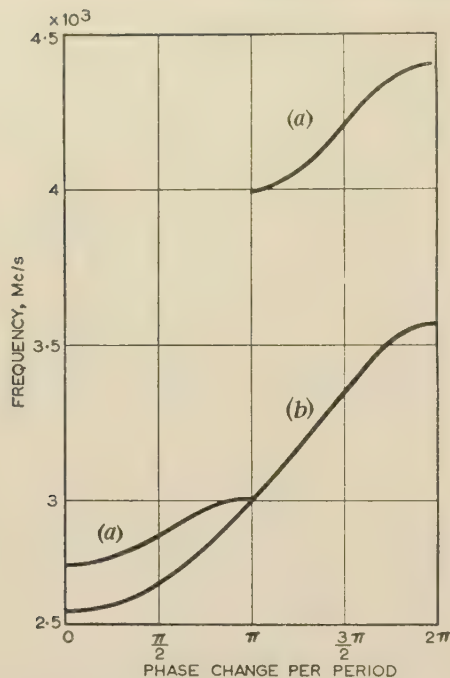
Fig. 3.—Dispersion curves for disc-loaded waveguides.

(a) With unmatched discs.
(b) With matched discs.

In (a) the waveguide diameter is slightly greater than required for matching. For comparison the two guides were made to have the same π_1-mode frequency by using slightly thinner discs in (a) than in (b).
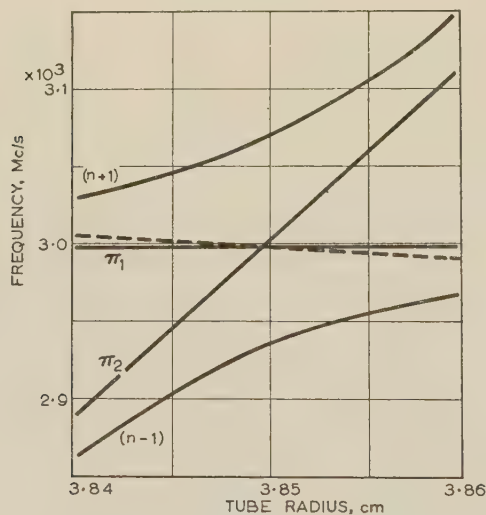
Fig. 4.—Resonant modes of a dielectric-disc-loaded cavity.

The cavity has ten sections each of length 5 cm, and the relative permittivity of the discs is 86. The disc thickness has been adjusted to keep the frequency of the π_1-mode fixed. The broken line gives the frequency at which the matching condition is satisfied.

π_1-mode, Fig. 4 shows that its nearest neighbour depends on which side of the matching frequency, f_m, it lies. When its frequency is less than f_m it lies on the upper pass-band and the $(n + 1)$-mode is its nearest neighbour, and when it is greater than f_m it lies on the lower pass-band and $(n - 1)$ is its neighbour.

Fig. 5 shows the mode separation over a wider range as a function of tube diameter for a particular 20-section cavity. The factor of improvement in frequency separation at the critical matching diameter, over the separation for a tube diameter only 3% different, is as much as 20 in this example.
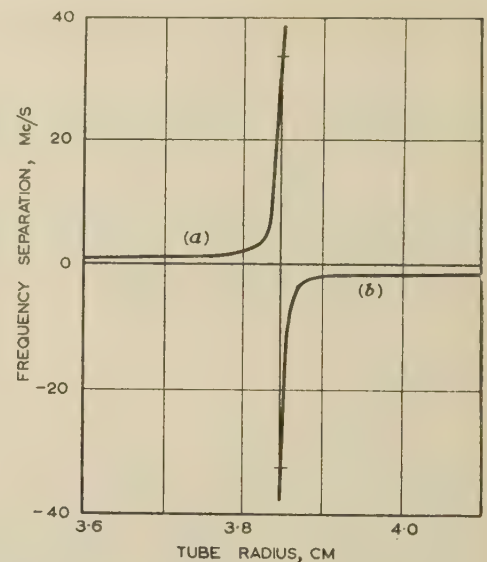
Fig. 5.—Mode separation for a 20-section cavity.

π_1-mode resonant frequency = 2 998 Mc/s.
$L = 5$ cm.
$\epsilon = 86$.

(a) Frequency separation between the modes = $n + 1$ and π_1.
(b) Frequency separation between the modes = $n - 1$ and π_1.
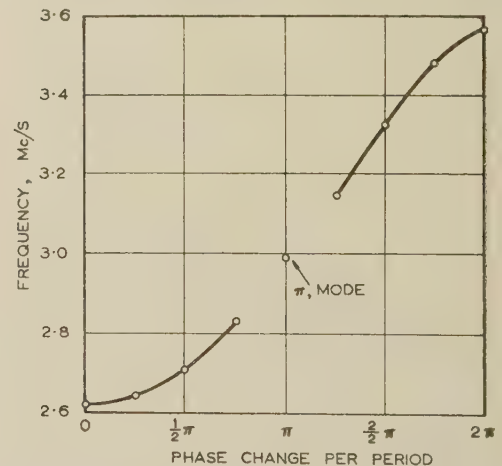
Fig. 6.—Measured confluent characteristic of a 4-section cavity.

$a = 3 \cdot 849$ cm.
$L = 5$ cm.
$2L_d = 0 \cdot 48$ cm.
$\epsilon \simeq 90$.

An experimental verification of a loaded cavity producing confluent band-pass characteristic was made, and the result shown in Fig. 6. A 4-section circular cylindrical cavity was constructed loaded by discs of a titania ceramic with a dielectric constant of about 90. The cavity diameter was chosen so that the matching condition held at 2 998 Mc/s, and for a periodic length of 5 cm the disc thickness was chosen so that this was also the π-mode frequency. The curve was plotted from the measured values of the resonant frequencies of the cavity. The cavity termination did not permit the π_2-mode to be excited, but from the plotted points it can be seen that if a finite stop band does exist, it must be extremely narrow.

It is unlikely that the matching condition was exactly satisfied as the dielectric constant was not known to within an accuracy of better than 2%, but cavity losses will tend to modify the curve in the vicinity of the π-mode to produce confluence if the theoretical stop band for the lossless cavity is small.

CAVITY LOADED WITH DIELECTRIC DISCS HAVING AXIAL HOLES

(4.1) Introduction

In practical applications of dielectric disc loading, it is necessary for there to be a concentric hole in each disc to allow for the passage of the particle beam. The hole perturbs the electromagnetic field pattern and makes a simple impedance match between the air and dielectric regions no longer possible. A complete mathematical treatment of this problem is very difficult, but the most important point to be established is whether, using discs with holes, it is still possible to obtain the confluence previously described. It is probably more convincing if this can be demonstrated in a practical way.

(4.2) Discussion of Problem

It has already been shown that confluence can occur if the loading obstacles are reflectionless at the π-mode frequency. Some attempts were made to discover whether a frequency could be found at which a single disc with a hole was reflectionless. This was very difficult to do experimentally, since the required condition is near the cut-off frequency of the unloaded guide, and also a special standing-wave indicator for a circular waveguide was required. However, on further consideration, it became clear that this reflectionless state is not a necessary condition nor is it even relevant. For solid discs a single propagating mode only need be considered, and at the π-mode frequency in the confluent case no interaction takes place between adjacent discs, i.e. the discs considered individually are reflectionless. When a hole is made in the disc, evanescent modes are set up, and it cannot be assumed that these have completely died out at adjacent discs. To be valid, therefore, an experiment must take into account any such mutual interaction between discs. Rather than investigate the question of reflections in a long periodically-loaded waveguide terminated, say, in an absorbing load it was decided to approach the matter from a different point of view. As has been shown, there are, in general, two frequencies at which there is a π-phase change per section. To measure these frequencies a length of guide may be converted into a resonant cavity by placing end walls midway between the discs or through the centre of the discs. It may be noted that only by placing end walls in these positions can the system be reflected in the end walls in such a way that the field pattern in a short cavity is the same as in a long structure. The shortest possible representative cavity contains only one disc, which can be placed at the centre of the cavity, or alternatively half-thickness discs may be placed against each end wall [see Figs. 7(a)
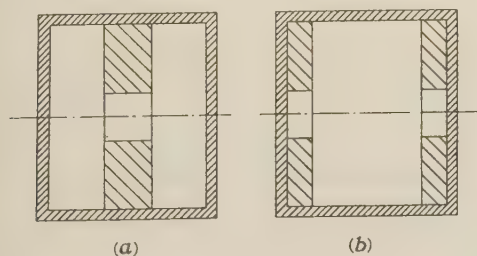
Fig. 7.—Disc positions for (a) π_1- and (b) π_2-mode cavities.

and 7(b)]. The resonant frequency of the cavity with the central disc gives one π-mode, and that of the cavity with the half disc gives the other π-mode. Using the notation of Section 3.3 these may be referred to respectively as the π_1- and π_2-modes. Thus by varying any dimension (such as the length or diameter of the cavity, the hole size or disc thickness), it can be shown

that it is possible for the π_1- and π_2-modes to have the same frequency, the condition for confluence has been established.

(4.3) Experimental Verification

An experimental verification was carried out. To save cost and for ease in machining, an initial test was done with discs of polystyrene instead of titania. As the dielectric constant is much less, a thicker disc was used so that the thickness would be roughly the same fraction of a wavelength. All factors are not altered in the same proportion by a change in dielectric constant, but there is no reason to suppose that there is any change in principle.

A cavity of length $5 \cdot 0$ cm and diameter $7 \cdot 24$ cm was prepared which had detachable end walls held in position by tie bars. The cavity was excited by a small probe placed in an axial hole in one end wall, the end of the probe lying flush with the wall. A similar probe in the other end wall was fed to a crystal detector, and resonance was detected in the manner of a transmission wavemeter, only a symmetrical E-mode being set up. The discs, which were a push fit in the cavity, were first located by a depth gauge and held in place by small screws through the guide wall. Resonant frequencies were measured for a range of disc thicknesses, the hole size being maintained at $\frac{1}{2}$ in diameter. The results are shown in Fig. 8. As the curves for the two π-modes
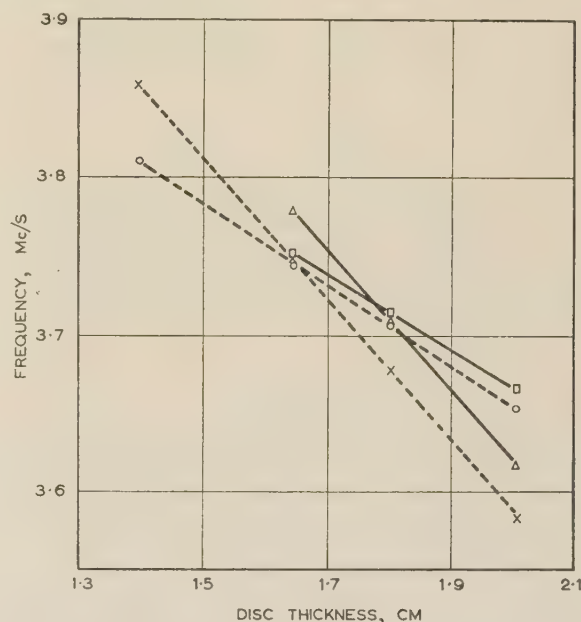
Fig. 8.—π-mode frequencies using polystyrene discs with and without holes.

$a = 3 \cdot 62$ cm.
$L = 5$ cm.
▢ π_1-mode, $0 \cdot 5$ in diameter hole.
○ π_1-mode, no hole.
△ π_2-mode, $0 \cdot 5$ in diameter hole.
✕ π_2-mode, no hole.

intersect when the disc thickness is $1 \cdot 77$ cm and the frequency is 3722 Mc/s, a confluent structure can be made. For comparison a similar test was carried out with solid polystyrene discs and the results are shown by the broken lines in the Figure. As expected, the effect of the hole is to increase the resonant frequency, but the general forms of the curves taken with and without the hole are similar. It is reasonable to suppose that any small change in dimensions will affect the resonant frequency of the solid-disc resonator in approximately the same way as for the resonator containing a disc with a hole, and so the simple calcula-

tions on solid discs may be used to estimate frequency shifts. This is particularly useful in reducing the number of measurements which require to be done to discover the precise cavity dimensions which give confluence at some particular frequency.

(5) THIN METAL DISCS

Most microwave accelerators in common use are loaded with thin metal irises, and operation of this type of accelerator at the π-mode is found to suffer from small mode separation. Slater[5], for example, reports a 0.1% mode separation at 10 cm wavelength for a 24-section cavity of the Massachusetts Institute of Technology resonant linear accelerator. It is not possible to improve this by trying to obtain confluence, as may be seen from the following argument.

In order that the π_1 and π_2 resonant frequencies shall be the same it is necessary for the loading obstacle to have the same effect on the resonant frequency of a cavity when placed in the middle, as when divided into two and placed against the end walls. If the loading obstacle is a thin metal disc (with or without a hole), then clearly when the half-discs are placed at the end walls the resonator behaves as if it were empty. For confluence, therefore, the disc when placed at the centre must not change the resonator frequency. This is not possible for a solid metal disc, and if it were possible to achieve the result by some arrangement of holes, clearly the disc no longer acts as a loading obstacle at that frequency. Hence it is impossible to achieve confluence in a slow-wave structure by the use of thin metal irises.

(6) ACKNOWLEDGMENTS

(7) REFERENCES

(1) R.-SHERSBY-HARVIE, R. B.: 'Theory of Dielectric Loaded Waveguides applied to Linear Electron Accelerators' (A.E.R.E. Report No. EL/R402; 1949).

(2) BELL, J. S., and WALKINSHAW, W.: 'Review of Theory of Dielectric Loaded Linear Accelerators' (A.E.R.E. Report No. G/R544; 1950).

(3) MULLET, L. B.: 'Dielectric Loaded Waveguides for Linear Accelerators at 10 cm. Wavelength' (A.E.R.E. Report No. G/R448; 1950).

(4) R.-SHERSBY-HARVIE, R. B.: 'A Proposed New Form of Dielectric-Loaded Waveguide for Linear Electron Accelerators', *Nature*, 1948, **162**, p. 890.

(5) SLATER, J. C.: 'Microwave Electronics' (Van Nostrand, 1950).

(6) FRY, D. W., and WALKINSHAW, W.: 'Linear Accelerators', *Reports on Progress in Physics*, 1948–49, **12**, p. 102.

(7) R.-SHERSBY-HARVIE, R. B.: 'Travelling Wave Linear Accelerators', *Proceedings of the Physical Society*, 1948, **61**, p. 255.

(8) LAWTON, E. J.: 'The Choice of Operating Mode for Standing Wave Type Linear Accelerators for Electrons', *Journal of Applied Physics*, 1948, **19**, p. 534.

(9) LAWTON, E. J., and HAHN, W. C.: 'Experimental Results on Standing Wave Type Linear Accelerators for Electrons', *ibid.*, p. 642.

(10) MULLET, L. B., CLAY, R. E., and HADDEN, R. J. B.: 'Multipactor Effect in Linear Accelerators and other Evacuated R.F. Systems, and a New Cold Cathode Valve' (A.E.R.E. Report GP/R1076).

(11) GUILLEMIN, E. A.: 'Communication Networks' (Wiley, 1935).

(12) LAMONT, H. R. L.: 'Theory of Resonance in Microwave Transmission Lines with Discontinuous Dielectric', *Philosophical Magazine*, 1940, **29**, p. 521.

(13) LAMONT, H. R. L.: 'The Use of the Waveguide for Measurement of Microwave Dielectric Constants', *ibid.*, 1940, **30**, p. 1.

(8) APPENDICES

Space Harmonics in Resonant Accelerator

Fig. 9 shows the type of dispersion curve which is associated with the space harmonics found in a periodically loaded accelerator. The forward harmonics, having positive group
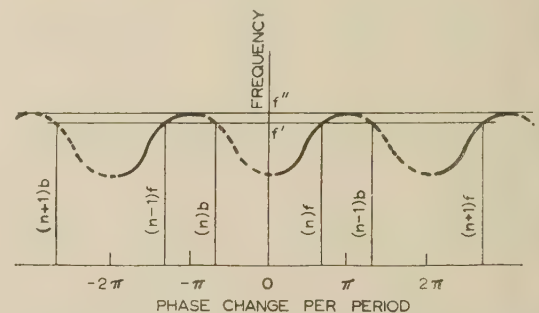
Fig. 9.—Space harmonics in a periodically loaded waveguide.

velocity, are shown as continuous lines, while the dotted lines show the backward harmonics. If the accelerator dimensions are chosen correctly, any one of the harmonics with positive phase velocity, i.e. $\beta_n L$ positive, can be made to travel at the same velocity as the particle and to resonate with it.

When the accelerator is operated as a resonant system, then for each forward travelling harmonic there is a backward harmonic of equal amplitude and of equal phase and group velocities, but of opposite sign. For example, if the accelerator is operated at the frequency f', the forward harmonics are $\ldots (n-1)_f$, $(n)_f$, $(n+1)_f$, etc., and the backward harmonics are $\ldots (n-1)_b$, $(n)_b$, $(n+1)_b$, etc. Of all these harmonics only one can be made to couple to the particle over a number of wavelengths and to give a resultant energy to it.

It is usual in accelerators to couple the particle to the first forward harmonic, $(n)_f$, and so this is designed to have the largest amplitude, such that most of the forward wave energy is carried in it. The other forward harmonics will have smaller and smaller amplitudes as their wavelengths decrease. Unavoidably in a resonant system, half the energy is still carried in the backward wave.

It is interesting to note the behaviour of the first backward harmonic with positive phase velocity, $(n-1)_b$, as the operating frequency approaches the upper cut-off frequency f''. It is found that its amplitude approaches that of the accelerating harmonic $(n)_f$, and as can be seen, its phase velocity also approaches that of $(n)_f$. However, so long as its phase velocity is slightly different from that of the particle, it can give no net acceleration to it. In the limit when the frequency f'' is reached, i.e. π-mode operation, the two harmonics coalesce. When the end walls lie midway
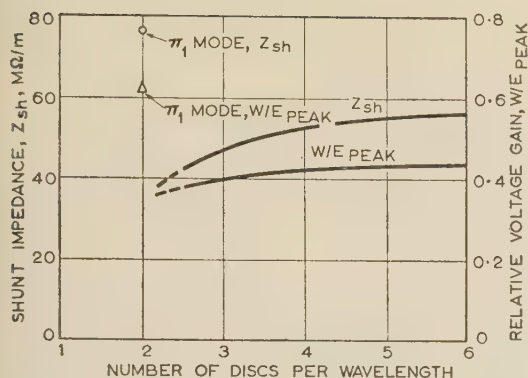
Fig. 10.—Performance of linear accelerator loaded with
dielectric discs.

$a = 3 \cdot 849$ cm.
$f = 2\,998$ Mc/s.
$\epsilon/\epsilon_0 = 86$.

tween discs, these harmonics add in phase and hence give an
:celerating field of greater amplitude.

If the end walls are moved a distance of one-quarter of the
harmonic wavelength, i.e. to points at the centre of discs, the
harmonics are out of phase so that they diminish the accelerating
field. The π-modes corresponding to these positionings of discs
are referred to in the paper as the π_1- and π_2-modes respectively.

To illustrate the advantage in operating a dielectric-disc-
loaded accelerator at the π-mode, calculations were made of
the energy gain of an electron travelling with approximately the
velocity of light in structures having different numbers of discs
per wavelength. Two discs per wavelength corresponds to the
π-mode, three discs to the $2\pi/3$-mode, etc. Fig. 10 shows a
graph of the relative voltage gain, defined as the ratio of the
mean energy gained per unit length, namely W electron-volts
per metre, to the peak field strength, E_{peak}. A graph has also
been made of the shunt impedance, defined as the ratio of W^2
to the mean power dissipated per unit length of the structure.
For ease in computation the case of solid discs was taken and
it was assumed that no energy was gained by the electron in
traversing a disc. The discontinuity at the π-mode is clearly
seen in both curves giving an improvement in both charac-
teristics by a factor of about two.

The advantage in operating a metal-iris-loaded accelerator at
the π-mode has been discussed by Fry and Walkinshaw.[6]

SYNCHRONIZED SYSTEMS WITH TIME DELAY IN THE LOOP

By Z. J. JELONEK, Dipl.Ing., Member, and C. I. COWAN, B.Sc., Graduate.

SUMMARY

Synchronized oscillators of the feedback type, possessing low-pass and band-pass filters in the loop, are considered. It is shown that these filters can, under certain conditions, be represented by time-delay networks.

Stability and the pulling effect in a synchronized oscillator with a time-delay network in the loop are investigated analytically, graphically and with the help of an analogue. The corresponding non-linear hystero-differential equation is studied, and typical solutions are presented on the phase plane. Methods of improving the performance by the use of correcting circuits are also considered.

LIST OF SYMBOLS

$M = \mu_b/S$ = Normalized detuning.
M_p = Critical value of M for pulling effect.
$p = \alpha + j\omega$.
$p \equiv d/dt$.
$p' = \alpha' + j\omega' = \alpha/S + j\omega/S$.
$p' \equiv d/d\tau$.
S = Half the angular synchronization range.
t = Time.
T = Time delay.
$T' = ST$ = Normalized time delay.
T_0 = Critical delay to cause instability.
T_p = Critical delay to cause pulling effect.
ϕ_1 = Instantaneous phase of the synchronizing signal.
ϕ_0 = Instantaneous phase of the controlled signal.
$\phi = \phi_0 - \phi_1$.
Λ = Limit cycle.
$\tau = St$ = Normalized time.
$\bar{\omega}_1$ = Angular frequency of the synchronizing signal.
ω_0 = Instantaneous angular frequency of the controlled signal.
$\bar{\omega}_0$ = Free-running angular frequency of the controlled signal.
$\mu_b = \bar{\omega}_0 - \bar{\omega}_1$ = Free-running beat angular frequency.

(1) INTRODUCTION

The system of synchronizing oscillators by indirect means, employing a phase discriminator and a reactance valve,[1] is widely used, its applications including frequency-generating circuits,[2] automatic frequency control,[3,4] and control of the horizontal sweep in television receivers.[5] The circuits used to perform the above operations generally include some means of filtering in the loop (see Fig. 1); these filtering networks increase the complexity of the system and affect the pulling effect,* transient and frequency response, and stability.

Although the differential equation describing such a system is non-linear, the analytical method can be used without great difficulty in problems regarding stability and frequency and

---

* The frequency range, in which an oscillator can be drawn into synchronism with all possible initial conditions, is called the pull-in range. Once synchronized, the oscillator can retain synchronism up to the limits of the pull-out range. The pulling effect is said to exist if the pull-in range is less than the pull-out range.

---

transient response within the synchronization range, since in the[se] cases a small perturbation within the system can be assume[d,] leading to a linear equation. On the other hand, in the probl[em] of pulling effect, the non-linear equation has to be consider[ed,] and, in this case, graphical analysis seems to give the best resul[ts.] The system which has attracted greatest attention is one with [a] simple RC low-pass filter following the phase discriminator, whe[re] the equation is of the type

$$A\frac{d^2\phi}{dt^2} + \frac{d\phi}{dt} + \cos\phi = B \quad . \quad . \quad . \quad [1]$$

A and B being constants.

This system is always stable. In some systems, howev[er,] which include more complex filters of the low-pass or band-pa[ss] type, such as those used for frequency generation, it may [be] necessary to consider not only the pulling effect but also t[he] stability of the system. For the purpose of analysis, it may [be] possible, under certain circumstances, to approximate to t[he] filter characteristics, by considering them as time-delay networ[ks] and establishing a system equation on this basis. This approa[ch] gives a non-linear difference-differential or hystero-differenti[al] equation of the first order, equivalent to a differential equation [of] infinite order. Some linear hystero-differential equations ha[ve] been analysed.[7-12] The present non-linear equation, howeve[r,] does not appear to be tractable analytically, but an analysis [of] a linearized equation gave conditions of stability and usef[ul] indications of the graphical procedure for solution. In additio[n,] observation of an experimental set using an electro-mechanic[al] analogue with a time-delay unit helped considerably in antic[i-]pating and checking graphical computations.

(2) DERIVATION OF THE HYSTERO-DIFFERENTIAL EQUATION

The analysis of a simple synchronizing system possessing [a] filter of transfer function $F(p)$ is carried out in order to devel[op] the phase equation.
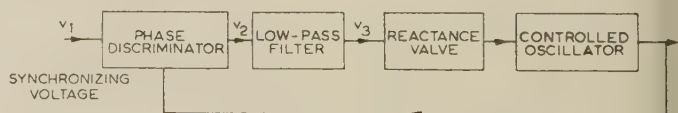
Fig. 1.—Block diagram of a synchronized system.

Referring to Fig. 1, put

$$\omega_0 = \bar{\omega}_0 - rv_3 \quad . \quad . \quad . \quad . \quad [$$

where r is the reactance valve constant (rad/sec-volt), and ω[0] and $\bar{\omega}_0$ are the instantaneous and free-running angular frequenci[es] of the controlled oscillator, respectively.

But
$$v_3 = F(p)v_2 \quad . \quad . \quad . \quad . \quad [$$

and
$$v_2 = q\cos\phi \quad . \quad . \quad . \quad . \quad [$$

where q is the discriminator constant (volt/rad), and ϕ is th[e]

ference between the phase of the oscillator voltage, v_0, and that of the synchronizing voltage, v_1.

Therefore

$$\omega_0 = \bar{\omega}_0 - qr\mathrm{F}(p)\cos\phi$$

$$\omega_0 - \bar{\omega}_1 = \bar{\omega}_0 - \bar{\omega}_1 - qr\mathrm{F}(p)\cos\phi$$

where $\bar{\omega}_1$ is the angular frequency of v_1.

So that

$$\frac{d\phi_0}{dt} - \frac{d\phi_1}{dt} = \frac{d\phi}{dt} = p\phi = \mu_b - qr\mathrm{F}(p)\cos\phi$$

where $\mu_b = \bar{\omega}_0 - \bar{\omega}_1$ is the undisturbed beat angular frequency. It can be seen that, in synchronism ($p\phi = 0$), μ_b can be within range $\pm qr\mathrm{F}(0)$, so that, assuming that $\mathrm{F}(0) = 1$, half of the pull-out range, or synchronization range, is

$$S = qr \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The phase equation then becomes[4]

$$p\phi + S\mathrm{F}(p)\cos\phi = \mu_b$$

in normalized form,

$$\frac{1}{S}p\phi + \mathrm{F}(p)\cos\phi = \frac{\mu_b}{S}$$

$$p'\phi + \mathrm{F}(p)\cos\phi = M \quad . \quad . \quad . \quad . \quad (6)$$

It is shown in Section 9.2 that if the network possesses a substantially flat amplitude response and linear phase response within a band of $2 \cdot 5S$ radians per second, then the low-pass filter in the synchronization circuit shown in Fig. 1 could be replaced by a time delay

$$T = -\frac{\Delta\theta}{\Delta\omega} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

without materially affecting the performance of the circuit. Thus, for an input $v_2(t)$, the output will be $v_3(t) = v_2(t - T)$.

However,

$$v_2(t - T) = v_2(t) - Tpv_2(t) + \frac{T^2}{2!}p^2v_2(t) - \ldots \quad (8)$$

$$= \varepsilon^{-Tp}v_2(t) \quad . \quad . \quad . \quad . \quad . \quad (8a)$$

Hence, the transfer function can be reduced to ε^{-Tp} or, in the normalized form,

$$\mathrm{F}(p) \equiv \varepsilon^{-T'p'} \quad . \quad . \quad . \quad . \quad (9)$$

where $T' = ST$, $p' = p/S \equiv d/d\tau$, and $\tau = St$ is the normalized one.

Substituting in eqn. (6) gives

$$\frac{d\phi}{d\tau} + \varepsilon^{-T'p'}\cos\phi = M \quad . \quad . \quad . \quad (10)$$

which could be written

$$\frac{d\phi}{d\tau} + \cos\phi(\tau - T') = M \quad . \quad . \quad (10a)$$

indicating that $\cos\phi$ is delayed but not affected otherwise. A similar approach can be made in the case of carrier systems utilizing band-pass filters and frequency changers which may be used for frequency generation (see Section 9.3).

Eqn. (10a) is called a non-linear hystero-differential equation, and, since $\varepsilon^{-T'p'}$ can be expanded into an infinite series, yields a differential equation of infinite order. The first part of the paper is concerned with the solution of the equation from the

points of view of stability and pulling effect. In the second part, the problem of stabilization of unstable systems is considered. It does not seem possible to solve this equation as it stands, and graphical methods have been applied.

(3) STABILITY AND PULLING EFFECT IN THE SYSTEM

(3.1) Description of Solutions

In this Section the types of solution of the governing eqn. (10a) are investigated on the phase plane. It is intended at this stage not to explain the method of graphical construction of the solution, which will be given in detail later, but rather to outline how it is affected by time delay.

When time delay is not present, i.e. $T' = 0$, the system equation is simply

$$\frac{d\phi}{d\tau} = M - \cos\phi \quad . \quad . \quad . \quad . \quad (11)$$

and this can be represented on the phase plane by σ, as shown in Fig. 2(a) for $M < 1$. The motion of a point in the plane is
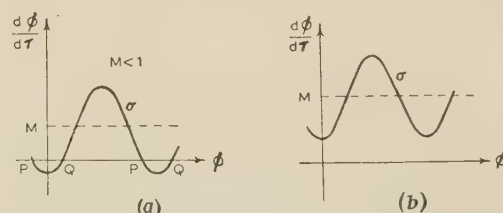
Fig. 2.—Simple phase equation on the phase plane.
(a) $M < 1$ (b) $M > 1$.

constrained to lie on this curve, moving from left to right above the ϕ-axis, and from right to left below the ϕ-axis, where $d\phi/d\tau$ is negative. The curve crosses this axis at the singular points P and Q where the solution can remain indefinitely, being stable at P and unstable at Q.

Thus, when $M < 1$, i.e. within the synchronization range, the system will eventually come to rest at a stable point P representing constant phase difference between the oscillators.

If, however, $M > 1$ [Fig. 2(b)], i.e. outside the synchronization range, the path will not cross the ϕ-axis, since $M - \cos\phi$ cannot be zero. Thus no locking is possible and beats will occur, with the phase difference between the oscillators increasing indefinitely.

When time delay is present, i.e. $T' > 0$, the point will move along a different path, which will lie, however, with the exception of the first period of time T when it can be arbitrary, between the limits

$$M - 1 \leqslant \frac{d\phi}{d\tau} \leqslant M + 1$$

The actual shape of the path is a function of T' and M. Thus, consider M to be reduced gradually from greater than unity to zero, while the delay is constant.

(a) $M > 1$. The system is operating outside the synchronization range, this state being represented on the phase plane by a periodic curve along which the operating point will move.

(b) $M_p < M < 1$. The system is now within the synchronization range, where locking may be expected to occur. Since, however, M is greater than a critical value M_p, which is dependent on T', the operating point continues to move along a periodic curve, and synchronism is not possible, although stable singular points have appeared. This curve is denoted by Λ_2 in Fig. 3, and corresponds to the limit cycle of the second kind, described in Reference 1.
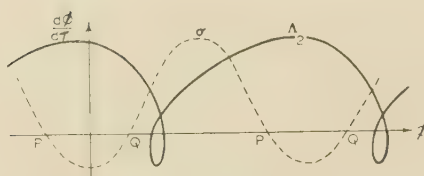
Fig. 3.—A periodic solution of eqn. (10a) on the phase plane.

(c) $0 < M < M_p$. Synchronism is now achieved, and Λ_2 has vanished. If $M < M_p$, with any initial conditions, the operating point tends either to a stable equilibrium at P, Fig. 4(a), or to move along another closed curve, around P, Fig. 4(b). This

(a)

(b)

(c)
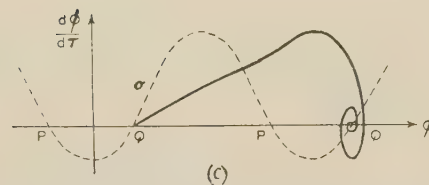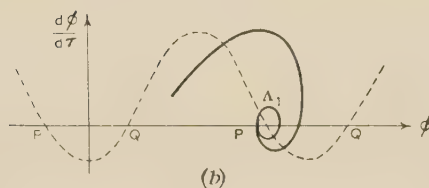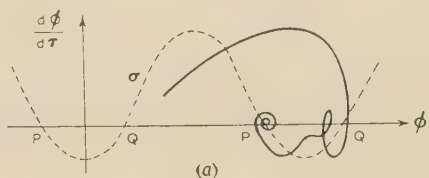
Fig. 4.—Examples of solutions of eqn. (10a) on the phase plane.

atter case represents instability at P, and its presence depends on the magnitude of T'. The curve denoted by Λ_1 corresponds to a limit cycle of the first kind.

(d) $M = M_p$. At the critical value of M, the operating point, if started from Q, comes to rest at the next point Q, Fig. 4(c). In a manner corresponding to the above, if M is held constant at less than unity, and T varied from a large value to zero, the pattern of behaviour is similar, provided that the point is originally on Λ_2. When T has been reduced to the critical value T_p (for a given M), Λ_2 vanishes.

The pulling effect exists, therefore, because, as M is being reduced, if the operating point moves along the limit cycle V_2, it will continue to do so as long as the latter exists. Thus the system 'pulls-in' to synchronism when $M = M_p$ (for a particular value of T'). Conversely, if the operating point is at a stable singular point, it will remain there until the latter vanishes. Thus, as M slowly increases, the system 'pulls-out' of synchronism when $M > 1$.

It is also possible, however, for the system to be locked even in the presence of Λ_2, provided that $|M| < 1$, if it starts from certain initial conditions. This aspect of performance will not be further considered.

(3.2) Stability

In order to consider the stability of the system at the sing[le] point P or Q, where the phase ϕ_2 satisfies $\cos \phi_2 = M$, a s[mall] increment x is added to ϕ in eqn. (10a):

$$\phi = \phi_2 + x$$

Then eqn. (10a) becomes

$$\frac{dx(\tau)}{d\tau} + \cos \phi_2 \cos x \,(\tau - T') - \sin \phi_2 \sin x \,(\tau - T') = M$$

Since x is very small, and $\cos \phi_2 = M$,

$$\frac{dx(\tau)}{d\tau} - \sin \phi_2 x(\tau - T') = 0 \quad . \quad . \quad . \quad .$$

This is a linearized form of the equation. It can be noticed [th]at point Q, $\sin \phi_2 > 0$, whereas at point P, $\sin \phi_2 < 0$. Section 9.1, which contains an analysis of the linear hyst[eresis] differential equation, the question of stability is examined. [It] is shown that, whereas point Q is always unstable, it is poss[ible] for point P to be stable or unstable, depending on the value o[f] The condition of stability is

$$-\tfrac{1}{2}\pi < T' \sin \phi_2 < 0 \quad . \quad . \quad . \quad .$$

In this Section the stability is considered in a different man[ner] to give an indication of the means of stabilization of an unst[able] system.

In order to do this, the system is considered at first with[out] time delay in the loop. Subsequently the effect of the dela[y is] considered as a phase shift increasing linearly with frequen[cy,] which must be added to the phase shift of the system.

Consider a system without time delay, in which the oscilla[tor] is operating in stable synchronism with the constant ph[ase] difference ϕ_2.

Then, referring to Fig. 1,

$$v_3 = v_2 = q \cos \phi_2$$

Consider now the loop opened between the phase discrimina[tor] and the reactance valve, a voltage $q \cos \phi_2$ being put into [the] reactance valve in order to maintain synchronism. If now [a] very small increment $V_3 \varepsilon^{j\omega t}$ is added to $q \cos \phi_2$,

$$v_3 = q \cos \phi_2 + V_3 \varepsilon^{j\omega t}$$

a component $V_2 \varepsilon^{j\omega t}$ will appear at the output of the phase [dis]criminator. The ratio V_2/V_3 will then be the open-loop trans[fer] function of the system.

The frequency of the oscillator will have an increment $rV_3 \varepsilon$[$^{j\omega t}$,] and the phase increment

$$rV_3 \int \varepsilon^{j\omega t} dt = \frac{rV_3}{j\omega} \varepsilon^{j\omega t}$$

so that

$$v_2 = q \cos \left(\phi_2 + \frac{rV_3}{j\omega} \varepsilon^{j\omega t} \right) \simeq q \cos \phi_2 - q \sin \phi_2 \frac{rV_3}{j\omega} \varepsilon^{j\omega t}$$

since $V_3 \ll 1$.

Hence, the open-loop transfer function is given by

$$K(j\omega) = -\frac{qr \sin \phi_2}{j\omega}$$

Since $qr = S$, and at P, ϕ_2 is negative,

$$K(j\omega) = \frac{S\sqrt{(1 - M^2)}}{j\omega} \quad . \quad . \quad . \quad ($$

$K(j\omega)$ has a polar plot as shown in Fig. 5(a). The integrati[on]
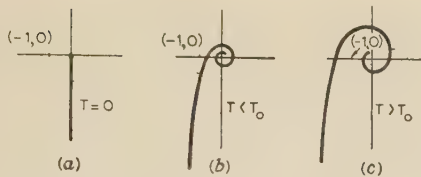
Fig. 5.—Polar plot of the open-loop transfer function.

$-\frac{1}{2}\pi$ phase shift, is the result of the reactance-valve/phase-discriminator arrangement.

As stated earlier, the effect of time delay is one of phase lag proportional to frequency, leading to a spiral in the polar diagram.

Figs. 5(b) and 5(c) show two cases of systems with delay. Since, in the second of these, the diagram encloses the point (1, 0), the system is unstable.

The condition of stability is that, at a frequency ω_1 at which $1 = 1$, i.e. at

$$\omega_1 = S\sqrt{(1 - M^2)} \quad . \quad . \quad . \quad . \quad (15)$$

the critical time delay must contribute an additional phase lag $\frac{1}{2}\pi$.

Thus

$$\omega_1 T_0 = \frac{1}{2}\pi$$

and

$$T_0 = \frac{\pi}{2S\sqrt{(1 - M^2)}}$$

$$T_0' = ST_0 = \frac{\pi}{2\sqrt{(1 - M^2)}} . \quad . \quad . \quad (16)^*$$

This agrees with the condition obtained in Section 9.1. The dependence of T_0' on M is shown in Fig. 6.
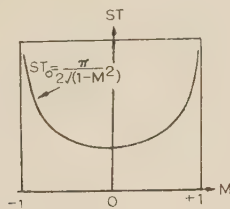
Fig. 6.—Limit of stability of the synchronized system.

The above approach suggests means of stabilizing unstable systems in a manner somewhat similar to that used in the case of servo mechanisms. This will be considered in a later Section.

(3.3) Pulling Effect

From observation of an electro-mechanical analogue of the above system, it is found that as time delay is increased, the pull-in range decreases, although the pull-out range is unaltered. Thus the pulling effect is dependent on delay, and it is desirable to find the relationship between them.

A simple expression for the numerical value of the pulling effect can be given as

$$\text{Pulling effect} = \frac{\text{(Pull-out range)} - \text{(Pull-in range)}}{\text{Pull-out range}}$$

$$= 1 - M_p . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

* Since the paper was submitted, an approximate analysis of synchronized systems has been published by Kapranov,[18] in which a triangular function of ϕ was substituted for $\cos \phi$. His results of the analysis of a system with an RC filter were found in reasonable agreement with those given in Reference 1. He also considered a system with a time delay, but the conclusions appear to be faulty because of an error in the analysis, owing to which the order of the differential equation obtained was not raised by the presence of the time delay. The condition of stability, however, is obviously in agreement with eqn. (16) with $M = 0$.

(4) GRAPHICAL AND EXPERIMENTAL INVESTIGATION OF THE PULLING EFFECT

Since an analytical approach seems to yield no solution of the pulling problem, graphical methods and an analogue were employed.

The observation of the operation of the analogue led to the conclusion that, at the edge of the pull-in range, the corresponding curve in the phase plane would move from one unstable point Q to the next one, as it was found to be the case in the system with an RC filter following the phase discriminator.[1]

In the graphical method, therefore, given the values of M and T, a curve was obtained in the vicinity of a point Q and continued into the neighbourhood of the following point Q. If it did not reach it, the value of T was modified and the procedure repeated.

(4.1) Analytical Basis of the Graphical Construction

In order to proceed with the graphical method, it is necessary first to know how the operating point will travel immediately after it has been displaced by an infinitesimal amount from Q. This information was obtained by a study of the linearized system equation (12). As found in Section 9.1, the general solution near Q is

$$x = A_1 \varepsilon^{\alpha_1' \tau} + \sum_{3}^{\infty} B_k \cos (\omega_k' \tau + \psi_k) \varepsilon^{\alpha_k' \tau}$$

where $\alpha_1' > 0$ and $\alpha_k' < 0$.

The starting impulse may give rise to all the above terms, but after a very short time has elapsed the oscillatory terms die out and the exponential term alone remains.

Therefore the operating point moves according to the law

$$x = A_1 \varepsilon^{\alpha_1' \tau} \quad . \quad . \quad . \quad . \quad . \quad (18)$$

giving

$$\frac{dx}{d\tau} = \frac{d\phi}{d\tau} = \alpha_1' A_1 \varepsilon^{\alpha_1' \tau}$$

A transcendental equation,

$$\alpha_1' = \sin \phi_2 \varepsilon^{-\alpha_1' T'} \quad . \quad . \quad . \quad . \quad (19)$$

is also derived, from which α_1' can be found.

Thus, the operating point will move with this gradient after the decay of the oscillatory terms, as shown in Fig. 7. This applies only in the region close to Q.
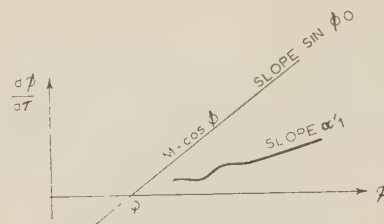
Fig. 7.—Movement of the operating point in the vicinity of point Q.

At the other end of the path, where the operating point approaches the next point Q, the approach may be such that the value of A_1 may be zero. Only then could the path lead finally to the point Q. This would be the case when, with a given M, T has a critical value equal to T_p.

The point would approach Q in a logarithmically decreasing spiral, since in the end only one oscillatory term, which attenuates at the slowest rate, would remain.

If $T \neq T_p$, the exponential term will exist and, depending on the sign of A_1, the point will move as, for example, in Fig. 8(a) or 8(b).

(a)

(b)
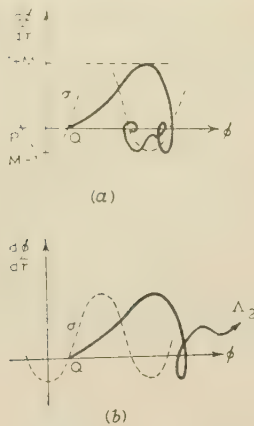
Fig. 8.—Solutions of eqn. (10a) on the phase plane.

(a) $T' < T'_p$
(b) $T' > T'_p$

(4.2) The Graphical Method*

Consider Fig. 9, which shows the phase plane with a plot, σ, of the simple phase equation (11). With time delay, the operating point moves along a path which must always satisfy eqn. (10a). As shown in Section 4.1, the point starts from Q at a gradient
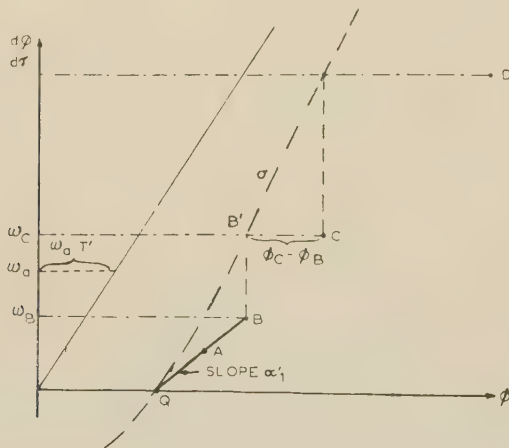
Fig. 9.—The construction of a solution on the phase plane.

α'_1. Let A and B be two points on this path close to Q, spaced such that the time taken to travel from A to B is equal to one period of time delay, T'. (The distance between A and B can be found by applying eqn. (18) with T' in place of τ.) Since eqn. (10a) must also be satisfied at B,

$$\omega_B = \frac{d\phi_B}{d\tau} = M - \cos \phi_A$$

This equation forms the basis for finding further points of the path.

Proceeding from B, the operating point moves along the path shown, passing through points C, D, etc., spaced one period of time delay apart. Projecting vertically upwards from B will give ω_C at the point B' of intersection with σ.

Since the velocity of the point moving from B to C increases from ω_B to ω_C in time T', it is necessary to find the average velocity ω_a lying between these two, so that

$$\phi_C - \phi_B = \omega_a T'$$

* A somewhat different method of graphical solution of this type of equation has been recently given in Reference 16.

It can be shown that, if the ratio between ω_B and ω_C is n[...] unity, the average velocity is very close to their mean value:

$$\omega_a \simeq \tfrac{1}{2}(\omega_B + \omega_C)$$

Thus, the co-ordinates of C can be found, and proceeding in [...] way the whole path taken by the operating point can [...] constructed.

If ω_C/ω_B differs greatly from unity, the overall accuracy [...] the method can be increased by dividing the time delay int[...] number of equal sub-periods, with a consequent decrease in [...] distance between successive points.

In this way, the point will move as shown in Fig. 8(a) or 8[...] according to whether T' is smaller or greater than T'_p. Thus, [...] trial and error, a satisfactory approximation to T'_p is found. [...] estimated error with which T'_p was found in this way was 1–2[...]

(4.3) Experimental Equipment and Methods

In the graphical construction outlined in Section 4.2 it [...] clearly advantageous if some approximation to the desired va[...] of T' can be obtained by other means directly, e.g. by the use [...] an experimental set-up or an analogue. This greatly reduces [...] number of graphical attempts at solution and serves also a[...] rough check. In addition, the observation of such an arran[...] ment allows for a better understanding of the effect of the var[...] tion of T' and M on operation of the system. In measuring [...] effect of stabilizing networks on the pull-in range, neith[...] analytical nor graphical methods can be readily employ[...] whereas the analogue gives results as conveniently as in [...] simple system.

In this investigation an analogue was employed whose functi[...] was to solve the hystero-differential equation (10a). It was bu[...] in two units: (i) the analogue of the simple phase equation (1[...] and (ii) a time-delay network.

(4.3.1) The Electro-Mechanical Analogue.[13]

The electro-mechanical analogue (see Fig. 10) consists of a d[...] motor-generator driven by an amplifier whose input is the su[...]
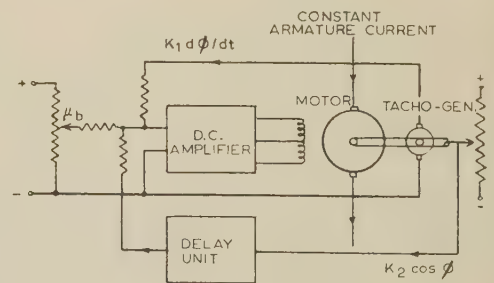
Fig. 10.—Analogue of the synchronized system with time delay.

of three voltages. One of these is obtained from a cosi[...] potentiometer, driven by the motor, giving an output volta[...] proportional to $S \cos \phi$. A voltage representing μ_b is deriv[...] from a hand-set potentiometer across a d.c. supply. The thi[...] voltage, obtained from the tacho-generator, is proportional [...] the speed of the motor, i.e. to $d\phi/dt$. The motion of the mot[...] is such that the sum of the three voltages is maintained close [...] zero level, thus satisfying eqn. (10a). The time scale and S c[...] be altered by changing the proportion of the generator volta[...] applied to the amplifier. This facility was very useful when t[...] analogue was used in conjunction with the delay unit.

(4.3.2) The Delay Unit.[14]

The basis of the delay system is a magnetic tape record[...] arranged so that the record and replay heads are in simultaneo[...]

peration. The length of delay is equal to the time taken for the
pe to travel from the former to the latter head, obviously a
nction of speed and head spacing.

Since the type of signal which must be delayed is of very low
equency or direct current, it is necessary to use some form of
odulation before recording. This is achieved by using a
equency-modulated oscillator to feed the recorder, which, after
production, delivers the signal to a discriminator of the double
ulse-counter type,[15] which recovers the modulating signal.

The delays obtained by this means were $0 \cdot 233$ sec and $0 \cdot 433$ sec,
ith two inter-head spacings, at a tape speed of $7\frac{1}{2}$ in/sec. The
ormalized delay $T' = ST$ could be varied continuously by
tering S, as explained in Section 4.3.1, within the range $0 \cdot 4$–$2 \cdot 5$.
By placing the delay unit in the path of the voltage feeding
ack from the cosine potentiometer to the amplifier input, an
alogue for eqn. (10a) was obtained.

The pulling effect could be measured by noting the difference
the hand-set dial readings for pulling-out and pulling-in for a
articular value of ST [see formula (17)]. The estimated accuracies
measurement of μ_b and ST were 5% and 3%, respectively.

(4.4) Results

The results for pulling effect, as obtained by graphical and
perimental means, are shown in Fig. 11, together with a plot
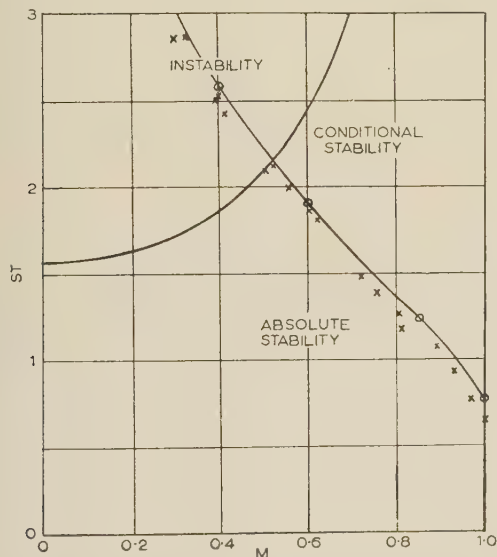
g. 11.—The effect of time delay on the pull-in range, together with
the limit of stability (eqn. 16).
    × Experimental.
    ○ Computed graphically.

the relation between T'_0 and M, eqn. (16), for stability. It is
be noted that the pulling effect starts at $T' = 0 \cdot 77$.

For a given value of T', the experimental results show some-
hat greater pulling effect. This is due, probably, to factors in
e mechanical construction of the analogue, e.g. friction and
acklash, the effect of the inertia of the motor, and the inter-
ence due to wow and flutter in the tape recorder.

The area of the graph can be divided into zones:

(a) Region of Absolute Stability.—In this area, the system
ways reaches synchronism, since it operates within its pull-in
ge and below the limit of instability.

(b) Region of Instability.—Within this area no stable syn-
ronism is possible, and the value of T' is such that a limit cycle
the first kind exists at P.

(c) Region of Conditional Stability.—The manner in which the
system will operate within the area is dependent on the direction
in which M varies.

(5) DESIGN OF STABILIZING NETWORKS

A designer of an automatic frequency-control system can find
in the previous Section sufficient data to predict stability and the
pulling effect in his circuit. If it appears unstable, he can
stabilize it by reducing S or by redesigning the filters with a view
to reducing T, or both. He may be restricted in these alterations,
however, and other methods of enlarging the stability area in
Fig. 11, i.e. extending the range of stability and reducing the
pulling effect, may become of interest.

Some simple stabilizing networks, as used in the servo-
mechanism technique, will be considsred here and their effect on
the pull-in range will be assessed.

(5.1) Phase-Lead Correction

Consider the phase-lead network shown in Fig. 12, whose
transfer function $K_1(p)$ is

$$K_1(p) = n_1 \frac{1 + pT_1}{1 + n_1 pT_1} \qquad \ldots \ldots \quad (20)$$
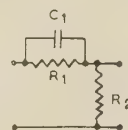
Fig. 12.—Phase-lead correcting network.

where $T_1 = R_1 C_1$ and n_1 is the attenuation at zero frequency;

i.e.
$$n_1 = \frac{R_2}{R_1 + R_2}$$

A relationship connecting the time-constant T_1 with the critical
delay for instability T_0 will now be developed. The open-loop
transfer function of the system, without the time delay, derived
in Section 3.2 becomes

$$K(p) = \frac{S\sqrt{(1 - M^2)}}{p} \qquad \ldots \ldots \quad (14a)$$

Therefore, the overall open-loop transfer function is

$$K_2(p) = \frac{S\sqrt{(1 - M^2)}}{p} \frac{1 + pT_1}{1 + n_1 pT_1} \qquad \ldots \quad (21)$$

if the attenuation n_1 of the correcting network is compensated by
an increase of gain of the system (this is necessary if the value of
S is to be preserved). Thus, from eqn. (21),

$$|K_2(j\omega)| = \frac{S\sqrt{(1 - M^2)}}{\omega} \frac{\sqrt{(1 + \omega^2 T_1^2)}}{\sqrt{(1 + n_1^2 \omega^2 T_1^2)}}$$

For instability, it is necessary that $|K_2(j\omega)| = 1$, which defines
the critical frequency ω_2. At this frequency the phase lead of
the network alone will be

$$\theta = \arctan \omega_2 T_1 - \arctan n_1 \omega_2 T_1$$

Now, for instability, the total phase shift must be π radians, and
since the overall phase shift of the system and the network is
$\frac{1}{2}\pi - \theta$, the critical time delay must retard the phase at this
frequency by $\frac{1}{2}\pi + \theta$. Thus

$$T_0 = \frac{\frac{1}{2}\pi + \theta}{\omega_2} \qquad \ldots \ldots \ldots \quad (22)$$

14
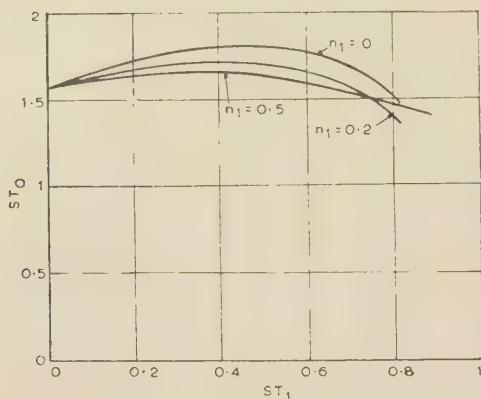
Fig. 13.—Range of stability of a synchronized system with phase-lead correction.

From the above formulae the normalized critical time delay ST_0 was found as a function of the normalized time-constant of the network ST_1, and is shown in Fig. 13 for a few values of n_1.

It is clearly shown in this graph that a phase-lead network can improve the system stability only within a certain range of ST_1. The maximum possible improvement is 16%, which occurs in the extreme case of $n_1 = 0$. It can also be seen that the use of a network having too large a time-constant can reduce stability.

The latter phenomenon can be explained with the help of Fig. 14, in which the amplitude and phase responses for an
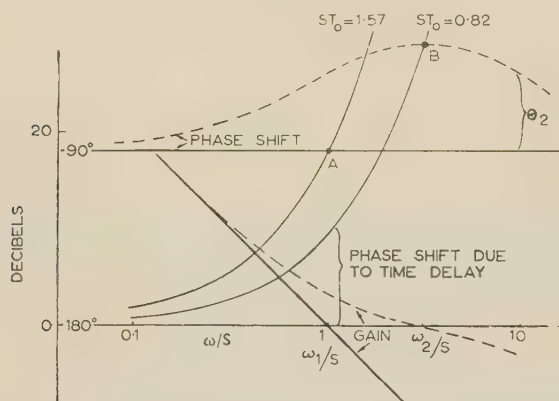
Fig. 14.—Gain and phase shift in a phase-lead corrected system.

$M = 0$, $ST_1 = 1$, $n_1 = 0 \cdot 1$.
—— Uncorrected system.
- - - - Corrected system.

uncorrected and corrected system without the time delay are plotted to a log-frequency scale, for $M = 0$, $ST_1 = 1$ and $n_1 = 0 \cdot 1$. In the uncorrected system (solid line) the phase response is $-90°$ at all frequencies. The critical time delay ST_0 must retard the phase by an additional $90°$ at a unit-gain frequency $\omega_1/S = 1$. Its phase lag is represented as the distance between the $-180°$ line and an exponential line passing through point A. In the corrected system (broken line) where the unit-gain frequency has been raised to ω_2/S, the new phase shift without the time delay is $\frac{1}{2}\pi - \theta_2$; hence, for instability, the time delay must contribute $\frac{1}{2}\pi + \theta_2$. The phase lag of the latter is represented by an exponential line passing through point B. In the present case, where the increase θ_2 does not offset the increase from ω_1 to ω_2 [see eqn. (22)], the critical time delay is less than that without correction, and so the correcting network fails to increase stability.

It was expected that this type of network would have another advantage, i.e. that it would reduce the pulling effect, since the

phase lead should in a way counteract the phase lag due to t[he] time delay. The result of tests carried out with a phase-le[ad] network ($ST_1 = 0 \cdot 5$, $n_1 = 0 \cdot 1$) are shown in Fig. 15, compar[ed]
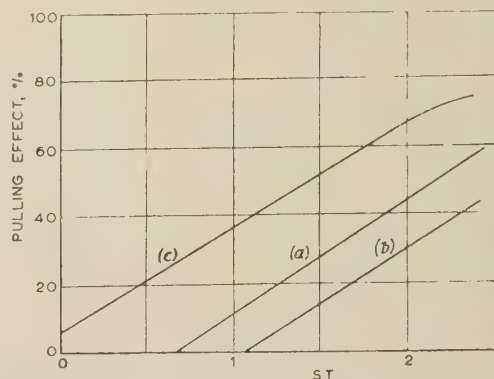
Fig. 15.—The pulling effect in corrected systems.
(a) No correction.
(b) Phase-lead correction.
(c) Phase-lead-lag correction.

with an uncorrected system. It can be seen that, as in the ca[se] of stability, the effect of the network is not very great.

It may be noted that the construction shown in Fig. 14 cou[ld] be used in designing correcting networks in servo mechanism[s] containing a time delay. In a conventional method of desig[n] the open-loop frequency responses are arranged so that the fr[e-] quency at which the phase is $-180°$ is higher than that at whi[ch] the gain is unity. In the case with a time delay, a similar criteri[on] would apply, in which an exponential line corresponding to t[he] time delay would replace the $-180°$ line.

(5.2) Phase-Lead/Phase-Lag Correction

It is apparent from the preceding Section that a great[er] stabilizing effect can be expected from a reduction of the un[it-] gain frequency than from advancing the phase. That reducti[on] could be obtained by using a phase-lag network. A system w[as] examined comprising the analogue, a phase-lag and a phase-le[ad] network (Fig. 16). The latter was added in order to compensa[te]
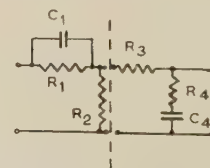
Fig. 16.—Phase-lead-lag correcting network.

for the phase lag of the former, in the appropriate frequen[cy] range. Fig. 17 shows the open-loop frequency response of th[e] system without a delay, for $ST_1 = 0 \cdot 63$, $n_1 = 0 \cdot 25$, ST_2 [=] $SR_4C_4 = 6 \cdot 3$, $n_2 = (R_3 + R_4)/R_4 = 4$. A critical time del[ay] of $ST_0 = 4 \cdot 6$ at ω_3 was obtained.

It should be mentioned that the above type of network allo[ws] for an almost unlimited improvement of stabilization of th[e] system. The great disadvantage, however, of this method [of] stabilization is that it adversely affects the pulling effect. Th[is] can be seen in Fig. 15, where the results of tests carried out wi[th] the analogue possessing the above networks are shown. It is [to] be expected that the use of the phase-lag network will increa[se] the pulling effect, from considerations in Reference 1.*

A specification of a synchronized system may require the pull[ing] range to have a certain value, without imposing any restriction [on]

* Although only stabilization and pulling in the synchronized systems are studi[ed] here, it was noticed that the phase-lag-lead network made the transient response ve[ry] sluggish.
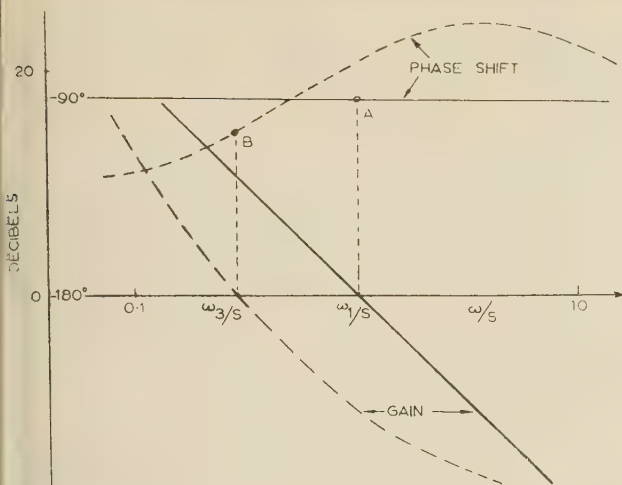
Fig. 17.—Gain and phase shift in a phase-lead-lag corrected system.
$M = 0$, $ST_1 = 0 \cdot 63$, $n_1 = 0 \cdot 25$, $ST_2 = 6 \cdot 3$, $n_2 = 4$.
——— Uncorrected system.
- - - - Corrected system.
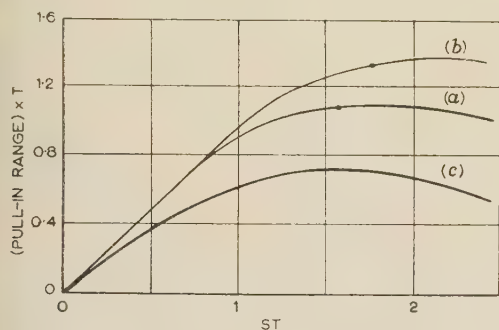
Fig. 18.—The pull-in range in corrected systems.
(a) No correction.
(b) Phase-lead correction.
(c) Phase-lead-lag correction.

pull-out range S. In many cases the latter value can be varied in the circuit with little difficulty. It may be of interest to the designer, therefore, to know how the pull-in range depends on the value of S, for a given fixed value of T. The relationships, deduced from the data in Fig. 15, and normalized with respect to T, are plotted in Fig. 18 for the three cases examined, i.e. a system with no correction, with phase-lead, and with phase-lag-lead correction. The points of instability are also marked on the first two curves. It can be seen that, whereas the phase-lead correction may increase the pull-in range by 15–22%, the phase-lag-lead correction only reduces it.

(6) CONCLUSIONS

(a) In synchronized oscillators of the feed-back type, low-pass and band-pass filters in the loop can be considered as time-delay networks, if the angular frequency range of approximately constant gain and linear phase shift is $2 \cdot 5S$ in low-pass filters and $5S$ in band-pass filters, S being half of the angular synchronization range.

(b) The presence of a time delay T in the loop of a synchronized oscillator may cause instability and introduce pulling effect, i.e. it can make the pull-in range smaller than the pull-out range.

(c) Equation (10a) of the system with time delay is of the first-order non-linear difference-differential or hystero-differential type, its solution in closed form is unknown as yet.

(d) In order to ensure the stability of the system, when it is within the synchronization range, T must be less than $\pi/2S$.

(e) In order to ensure that there is no pulling effect, T must be less than $0 \cdot 77/S$ approximately.

(f) A conventional phase-lead network incorporated into the system can improve the stability by allowing for an increase of the limit for T, as given in (d) above, but only by 16% at the most. From the point of view of the pulling effect, it can raise the limit for T, as given in (e) above, by 30–40%.

(g) A conventional phase-lead/phase-lag network can improve stability very considerably, but it may also greatly increase the pulling effect. It can even produce the pulling effect without any time delay in the system.

(8) REFERENCES

(1) JELONEK, Z., CELINSKI, O., and SYSKI, R.: 'Pulling Effect in Synchronized Systems', *Proceedings I.E.E.*, Monograph No. 79 R, November, 1953 (**101**, Part IV, p. 108).
(2) CAMFIELD, G. J.: 'A Frequency-Generating System for V.H.F. Communication Equipment', *ibid.*, Paper No. 1605 R, March, 1954 (**101**, Part III, p. 85).
(3) PRESTON, G. W., and TULLIER, J. C.: 'Lock-In Performance of an A.F.C. Circuit', *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 249.
(4) GRUEN, W. J.: 'Theory of A.F.C. Synchronization', *ibid.*, p. 1043.
(5) GEORGE, E.: 'Synchronizing Systems for Dot-Interlaced Colour Television', *ibid.*, 1951, **39**, p. 124.
(6) COWAN, C. I.: 'An Investigation of Stability and Pulling Effect in Synchronized Oscillators with Constant Time Delay', Post-graduate Diploma Thesis, Royal Technical College, Glasgow, 1955.
(7) CALLENDER, A., HARTREE, D. R., and PORTER, A.: 'Time-Lag in a Control System', *Philosophical Transactions of the Royal Society*, 1936, **235A**, p. 415.
(8) SILBERSTEIN, L.: 'On a Hystero-Differential Equation arising in a Probability Problem', *Philsophical Magazine*, 1940, **29**, p. 75.
(9) MINORSKY, N.: 'Self-Excited Oscillations in Dynamical Systems possessing Retarded Action', *Transactions of the American Society of Mechanical Engineers*, 1942, **64**, pp. A65 and A195.
(10) PIPES, L. A.: 'The Analysis of Retarded Control Systems', *Journal of Applied Physics*, 1948, **19**, p. 617.
(11) PALMER, P. J.: 'The Behaviour of Control Systems with Definite Time Delay and Random Disturbance', *Proceedings I.E.E.*, Monograph No. 73 M, August, 1953 (**101**, Part IV, p. 27).
(12) KUEPFMUELLER, K.: 'Die Forgaenge in Regelsysteme mit Laufzeit', *Archiv der Elektrischen Uebertragung*, 1953, **7**, p. 71.
(13) ANTOSZEWSKI, E.: 'Design of an Electro-Mechanical Analogue of a Synchronized Oscillator', Diploma Thesis, Polish University College, London, 1951.
(14) COWAN, C. I.: 'Time-Delay Network using a Magnetic Tape Recorder', Associateship Thesis, Royal Technical College, Glasgow, 1954.
(15) HUPERT, J. J., PRZEDPELSKI, A., and RINGER, K.: 'Double-Counter F.M. and A.F.C. Discriminator', *Electronics*, 1952, **25**, No. 12, p. 124.

(16) CUNNINGHAM, W. J.: 'Graphical Solution of Certain Non-Linear Differential-Difference Equations', *Journal of the Franklin Institute*, 1956, **261**, p. 621.

(17) CUNNINGHAM, W. J.: 'Non-Linear Oscillators with Constant Time Delay', *ibid.*, p. 495.

(18) KAPRANOV, M. V.: 'Pull-In Range in Phase-Controlled A.F.C. Systems', *Radiotechnika*, 1956, **11**, No. 12, p. 37.

(9) APPENDICES

(9.1) Analysis of a Linear Hystero-Differential Equation

(9.1.1) Characteristic Equation and its Roots.

Eqn. (12), developed in Section 3.2, describes the behaviour of a synchronized system in the immediate neighbourhood of singular points of equilibrium, P and Q. It can be written in a more convenient form:

$$\frac{dx(\tau)}{d\tau} - mx(\tau - T') = 0 \quad . \quad . \quad . \quad (12a)$$

where, referring to the parameters of the synchronized system, $m = \sin \phi_2 = +\sqrt{(1 - M^2)}$ at Q, and $m = -\sqrt{(1 - M^2)}$ at P, the normalized time $\tau = St$ and the normalized time delay $T' = ST$. An equation of this type has been studied[7-12] from various points of view, depending on the particular application. Here a more complete analysis will be presented. Although in the present application m is restricted to the range $-1 \leqslant m \leqslant 1$, the analysis will not be affected by this restriction.

Since eqn. (12a) can be considered as a linear equation of infinite order, its solution can be composed of terms such as $A\varepsilon^{p'\tau}$, where $p' = p/S$. Substituting into eqn. (12a) gives the characteristic equation

$$p' = m\varepsilon^{-p'T'} \quad . \quad . \quad . \quad . \quad (23)$$

which may have real and complex roots $p'_k = \alpha'_k + j\omega'_k$. It is convenient to normalize eqn. (23) further by putting

$$p'T' = \alpha'T' + j\omega'T' = \xi + j\eta$$

$$mT' = \mu$$

Then eqn. (23) becomes

$$\xi + j\eta = \mu\varepsilon^{-(\xi+j\eta)} \quad . \quad . \quad . \quad (23a)$$

If there are any real roots ξ_k, eqn. (23a) becomes

$$\xi = \mu\varepsilon^{-\xi} \quad . \quad . \quad . \quad . \quad (23b)$$

Referring to Fig. 19, which shows graphs of $\rho_1 = \xi/\mu$ and
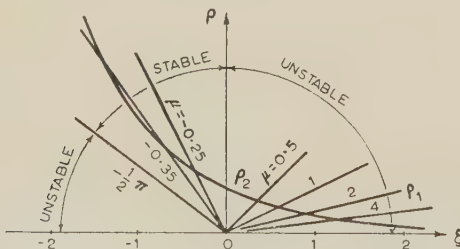
Fig. 19.—Construction for finding real roots of the characteristic equation (23).

$\rho_2 = \exp(-\xi)$ for various values of μ, it is apparent that there are four possible cases:*

(a) One real positive root. For all $\mu > 0$ the intersection of the graphs lies in the first quadrant.

(b) Two real negative roots. For $-1/\varepsilon < \mu < 0$ it is apparent that two intersections exist, such that $\xi_1 < -1 < \xi_2$.

---
* Real roots have been recently discussed in Reference 17 in the case of a more general hystero-differential equation.

(c) One double negative root. For $\mu = -1/\varepsilon$, $\xi_{1,2} = -1$.
(d) No real roots for $\mu < -1/\varepsilon$.

Thus, the general solution of eqn. (12a) can be put in the form

$$x = A_1\varepsilon^{\alpha'_1\tau} + A_2\varepsilon^{\alpha'_2\tau} + \sum_{k=3}^{\infty} B_k\varepsilon^{\alpha'_k\tau} \cos(\omega'_k\tau + \psi_k) \quad (24$$

This solution applies fully in case (b); whereas in case (a) $A_2 = 0$; in case (d), $A_1 = A_2 = 0$; and in case (c),

$$x = (A_1 + A_2\tau)\varepsilon^{\alpha'_{1,2}\tau} + \sum_{k=3}^{\infty} B_k\varepsilon^{\alpha'_k\tau} \cos(\omega'_k\tau + \psi_k) \quad (24a$$

In order to find the complex roots, eqn. (23a) is expanded into its real and imaginary parts:

$$\xi = \mu\varepsilon^{-\xi} \cos \eta$$

$$\eta = -\mu\varepsilon^{-\xi} \sin \eta \quad . \quad . \quad . \quad . \quad (25$$

Elimination of $\mu \exp(-\xi)$ from the above equations gives relation

$$\xi = -\eta \cot \eta \quad . \quad . \quad . \quad (26$$

in which the parameter μ does not appear. It is plotted i
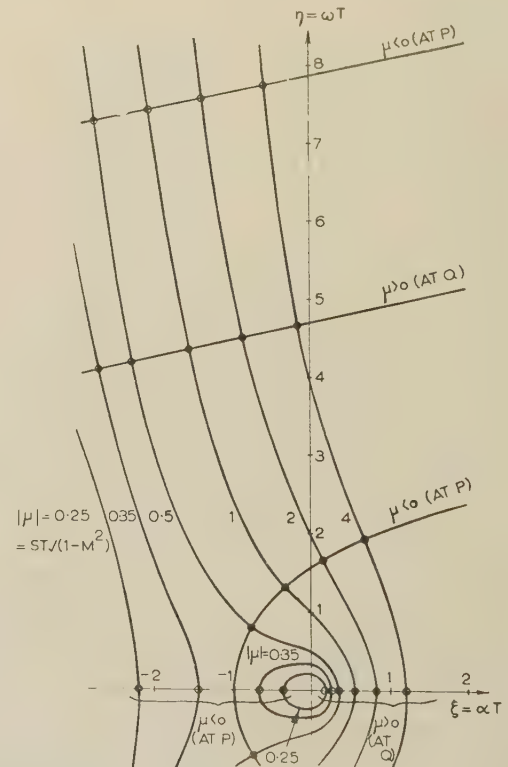
Fig. 20.—Roots of the characteristic equation (23a).

Fig. 20 with alternate branches corresponding to $\mu > 0$ o $\mu < 0$. Elimination of $\sin \eta$ and $\cos \eta$ from eqn. (25) leads t

$$\eta^2 = \mu^2\varepsilon^{-2\xi} - \xi^2 \quad . \quad . \quad . \quad . \quad (27$$

which gives a family of curves in Fig. 20, with a parameter μ Since all curves corresponding to eqns. (26) and (27) are sym metrical with respect to the axis $\eta = 0$, only one-half of the plan is shown. The intersection points of a curve of the family corresponding to a given μ, with the branches of the curve of eqn. (26), selected according to the sign of μ, give the comple roots of eqn. (23a). The intersection points of that curve wit

the axis $\eta = 0$, lying to the right of zero, give the real roots of eqn. (23a) if $\mu > 0$; those lying to the left of zero give the real roots if $\mu < 0$, in accordance with the previous analysis of eqn. (23b).

(9.1.2) Behaviour of the Synchronized System at the Points of Equilibrium.

The behaviour of the synchronized system will now be considered when it is at one of the states of equilibrium corresponding to points P or Q on the phase plane. First, it is to be noticed that the system is stable if all roots in Fig. 20 are to the left of the axis $\xi = 0$.

(a) At point Q, $\mu > 0$, there exists one real positive root and an infinity of complex ones. Thus, in the solution (24), $\alpha'_1 > 0$, $\alpha'_2 = 0$. It can be seen from Fig. 20 that $\alpha'_1 > \alpha'_3 > \alpha'_4 \cdots$ Substituting $\xi = 0$ in eqn. (25) gives pairs of values: $\pm \eta = \mu = 3\pi/2$, $\pm \eta = \mu = 7\pi/2$, etc. One can deduce from this and from Fig. 20 that, if $\mu < 3\pi/2$, all complex roots will have negative real parts, i.e. $0 > \alpha'_3 > \alpha'_4 \cdots$. If, however, $3\pi/2 < \mu < 7\pi/2$, one pair of complex roots will have a positive real part: $\alpha'_1 > \alpha'_3 > 0 > \alpha'_4 \cdots$. Theoretically, it is possible to find initial conditions such that only one or two terms in eqn. (3.3) will remain. These initial conditions must specify the state of the system during a period of time $-T < t < 0$.

(b) At point P, $\mu < 0$, several cases can be considered, according to the value of μ:

(i) $-1/\varepsilon < \mu < 0$. In this case there are two real roots, $\alpha'_2 < -1/T' < \alpha'_1$, and the remaining complex roots have a negative real part.

(ii) $\mu = -1/\varepsilon$. In this case there is a real double root $\alpha'_{1,2} = -1/T'$, and the system behaves according to eqn. (24a).

(iii) $-\pi/2 < \mu < -1/\varepsilon$. Now all roots are complex, with negative real parts. The above lower limit is found by substituting $\xi = 0$ in eqn. (25). Pairs of values are then obtained: $\pm \eta = -\mu = \pi/2$, $\pm \eta = -\mu = 5\pi/2$, etc. Hence, one pair of complex roots will have a negative real part if $-5\pi/2 < \mu < -\pi/2$. This gives the condition of stability of the system, i.e. $\mu > -\pi/2$, which, put in terms of the synchronized system, gives a relation obtained in a different way in Section 3.2:

$$ST < \frac{\pi}{2\sqrt{(1 - M^2)}} \quad . \quad . \quad . \quad . \quad (16)$$

(9.2) Transient Response and its Spectrum at the Critical Time Delay

The analyses of synchronized systems with time delay in the feedback loop can be applied with a fair amount of confidence to systems possessing low-pass or band-pass filters in the loop, provided that the latter satisfy certain requirements.

As far as stability is concerned, the gain around the loop should be constant and the phase shift linear for all values of $\omega_1 = S\sqrt{(1 - M^2)}$ (eqn. 15), i.e. from $\omega = 0$ to $\omega = S$, since ω_1, the frequency of oscillation of the system with the critical time delay T_0, is the only frequency entering the problem of stability. It corresponds to the lowest imaginary root for $\mu < 0$ in Fig. 20, and refers to the behaviour of the system in the immediate vicinity of point P in the phase plane.

The pulling effect is associated with the transient performance of the system having the critical time delay T_p: the system, in unstable equilibrium at point Q, when given an infinitesimal departure from equilibrium towards increasing ϕ [Fig. 4(c)] and positive $d\phi/dt$, should move to the next point Q and remain there; the final part of the movement is oscillatory, corresponding to the lowest root for $\mu > 0$ in Fig. 20. If the frequency of that root were taken as a measure of the required range of linearity of the filter characteristic, a value of $\omega = 4\cdot1S$ would be

obtained.* It will be shown, however, that this value is unnecessarily large. Transient responses have been obtained as a by-product of the graphical computation of T_p for two cases, $M = 1$ and $M = 0\cdot4$, and their frequency spectra were numerically computed. (Other cases, for intermediate values of M, may be expected to fall between these two.) It was found that over 90% of the area under each of the two curves was contained in the frequency band from $\omega = 0$ to $\omega = 2\cdot5S$. Therefore, this figure can be safely taken as the required range of linearity of the filter characteristic, for the filter to be replaced by time delay for the purpose of analysis.

(9.3) Systems with a Band-Pass Filter in the Feedback Path

Some frequency-generating circuits[2] use frequency changing in the path from the controlled oscillator to the phase discriminator, for the purpose of incorporating crystal-stabilized reference frequencies in the system. Frequency changers, of necessity, possess band-pass circuits in order to select the required component out of a multitude present in the output. Consider such a system, in which all frequency-selective circuits are lumped in one equivalent band-pass filter with a transfer function $F_2(p)$. In Fig. 1, this filter would be placed in the feedback path. In the system, the oscillator voltage v_0, which is frequency or phase modulated, is applied to the band-pass filter $F_2(p)$. The frequency or phase, ϕ_4, of its output is not a linear function of the modulating voltage v_3, since it suffers distortion due to the non-uniform gain and non-linear phase-shift of $F_2(p)$. Since the relationship between ϕ_4 and v_3 is very involved, it is easy to see that the presence of $F_2(p)$ in the system makes its analysis almost intractable.

The situation changes radically, however, if $F_2(p)$ has uniform gain K_2 and linear phase-shift θ_2 in a frequency band embracing the spectrum of v_0 in all cases of interest. $F_2(p)$ can then be considered to introduce gain K_2 and time delay $T_2 = -\Delta\theta_2/\Delta\omega$ into the system. In the derivation of the equation of the system, on the lines shown in Section 2, eqns. (3) and (4) would be replaced by

$$v_3 = F_1(p)q \cos \phi_4$$
$$= F_1(p)q \cos \phi(t - T_2)$$

If, in addition, $F_1(p)$ can be represented by a time delay T_1,

$$v_3 = q \cos \phi \, (t - T_1 - T_2) = q \cos \phi \, (t - T)$$

where the total time delay

$$T = T_1 + T_2$$

and eqn. (10a) and its subsequent analysis would apply to the system.

Consider now the requirement for the range of linearity of the characteristics of $F_2(p)$. The spectrum of a sinusoidally frequency-modulated carrier is somewhat wider than twice the highest modulating frequency, or the range of frequency deviation, whichever is greater. The greatest range of the frequency deviation of an oscillator which has to be within the synchronization range is $\pm2S$; this applies when the synchronizing frequency is constant and the oscillator is detuned; (with constant tuning of the oscillator and detuning of the synchronizing voltage, the range is $\pm S$). On the other hand, it is shown in Section 9.2, that the spectrum of v_3 is confined practically to a range from 0 to $2\cdot5S$, the components at frequencies greater than $2S$ being rather small. Hence, it may be concluded with some confidence that the required band of linearity of the band-pass filter $F_2(p)$ is $5S$.

---

* This corresponds to the case of $M = 1$, and is obtained from Figs. 11 and 20 in the following way: at $\mu = 0$, $\xi = -\infty$ and from eqn. (26), $\eta = \omega' T_p = \pi$. Since $T_p = 0\cdot77$, $\omega' = \pi/0\cdot77 \simeq 4\cdot1$. For other values of M smaller values of ω' are obtained.

A CONTRIBUTION TO THE DESIGN OF MULTI-ELEMENT DIRECTIONAL COUPLERS

By J. W. CROMPTON, B.E., Associate Member.

SUMMARY

A multi-element directional coupler is considered as a cascaded set of 2-element directional couplers, each of which, at the design frequency, is perfect in match and directivity. A characteristic impedance matrix and a voltage transfer matrix may be defined for each of the basic 2-element couplers, and these are analogous to the characteristic impedance and the propagation coefficient in iterative 4-terminal network theory.

The voltage transfer matrix is shown to be a linear function of the cosine and sine of a certain angle which is related in a simple way to the coupling factor of the elemental coupler. This function has algebraic properties analogous to those of complex numbers. In particular, the use of de Moivre's theorem is shown greatly to simplify the design of multi-element couplers.

An example is given of the design of a 5-slot 3 dB coupler (a binomial slot hybrid).

LIST OF SYMBOLS

λ_g = Wavelength in main waveguide.
c = Power coupling factor = $\sin^2 \theta$.
b' = Width of branch waveguide.
b = Width of main waveguide.
Z = Characteristic impedance of branch waveguide.
$Z_0 = \sqrt{(1 - Z)^2}$.
V_1', V_2', I_1', I_2' = Input voltages and currents to a 2-stub coupler.
V_1'', V_2'', I_1'', I_2'' = Output voltages and currents from a 2-stub coupler.
n = Number of 2-stub couplers in a semi-infinite chain.

(1) INTRODUCTION

The study and design of multi-element directional couplers may be approached in either of two ways. The first, which is the more general, takes as its starting-point the individual coupling elements—the holes or slots or other means used to couple the main transmission line to the auxiliary line. Unfortunately the analysis generally leads to expressions of great complexity which fail to demonstrate the underlying physical principles involved. Lippmann[1] has used this method to study the properties of couplers having two identical coupling elements, the analysis in this case yielding results which can be readily interpreted physically.

By using 2-element couplers as elements from which to build multi-element couplers, a second and less general study of their properties is possible. In this way Lippmann extended his work on the 2-element coupler to a complex analysis of a cascade set of 2-element couplers. The object of the paper is to present a simpler version which emphasizes the physical principles and enables the coupling factors and the dimensional tolerances of the elements to be easily calculated.

The discussion will be limited to waveguide couplers in whic[h] the coupling elements are $\lambda_g/4$ stubs connecting the broad face[s] of the main and auxiliary waveguides, but the principles used ar[e] quite general and may be applied to other types of coupler.

(2) THE 2-STUB COUPLER

The 2-stub coupler with ideal properties at the design frequenc[y] has been described by Lippmann. It is shown diagrammaticall[y] in Figs. 1(*a*) and 1(*b*) and has the following characteristics at th[e] design frequency:

(*a*) The power coupling factor [i.e. (power in auxiliary line/inp[ut] power)] is *c*.
(*b*) The directivity is infinite.
(*c*) The image impedances in both the main and auxiliary line[s] are equal.
(*d*) The fields at corresponding sections of the main and auxiliar[y] waveguides such as A and A′ in Fig. 1(*a*) are in phase quadrature.
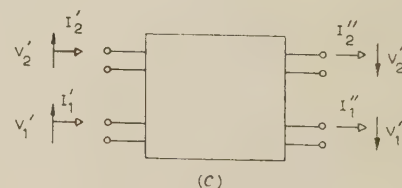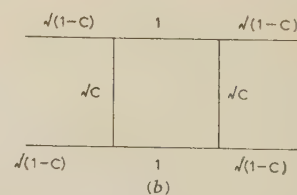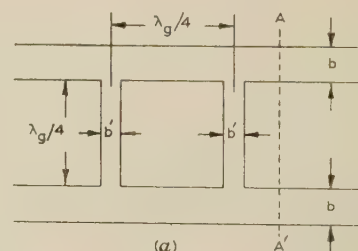
Fig. 1.—Configuration and characteristics of a 2-stub coupler.

(*a*) Mechanical configuration of a 2-stub coupler.
(*b*) Impedances in a 2-stub coupler.
    c = Power coupling factor for the coupler.
(*c*) Positive directions of voltage and current at the terminals of the 2-stub coupler[.]

The power coupling factor *c* is equal to the square of the rati[o] of the characteristic impedance of the stub guides to the main (o[r] the auxiliary) guide impedance;

i.e.

$$c = \left[\frac{b'}{b}\right]^2 . \quad . \quad . \quad . \quad . \quad . \quad (1$$

then the coupling is weak. In the case of stronger coupling, corrections to this simple relation are necessary to allow for the effect of the T-junctions at the ends of the stubs. These corrections are dealt with fully by Harrison.[2]

Corrections have also to be applied to the quarter-wavelength distances between the main and auxiliary waveguides and between the centre-lines of the coupling guides. Curves showing these corrections are given by Harrison.

(2.1) The Characteristic Impedance Matrix

The 2-stub coupler may be represented by an equivalent 8-terminal (4-terminal pair) network for which a characteristic impedance matrix and a voltage transfer matrix may be defined with properties analogous to those of the characteristic impedance and propagation coefficient in the case of iterative 4-terminal networks. Thus, using the symbols defined in Fig. 1(c) for the voltages and currents in the main and auxiliary waveguides, we may write for the input voltages in terms of the input currents when the coupler is correctly terminated

$$\begin{bmatrix} V_1' \\ V_2' \end{bmatrix} = \begin{bmatrix} Z_0 & 0 \\ 0 & Z_0 \end{bmatrix} \begin{bmatrix} I_1' \\ I_2' \end{bmatrix}, \quad Z_0 = \sqrt{(1 \quad Z^2)} \quad . \quad . \quad (2)$$

where Z = Effective characteristic impedance of the $\lambda_g/4$ stubs normalized with respect to the characteristic impedances of the parts of the main and auxiliary waveguides which lie between the stubs.

The output voltages and currents are similarly related:

$$\begin{bmatrix} V_1'' \\ V_2'' \end{bmatrix} = \begin{bmatrix} Z_0 & 0 \\ 0 & Z_0 \end{bmatrix} \begin{bmatrix} I_1'' \\ I_2'' \end{bmatrix}, \quad Z_0 = \sqrt{(1 - Z^2)} \quad . \quad . \quad (3)$$

The matrix

$$\begin{bmatrix} Z_0 & 0 \\ 0 & Z_0 \end{bmatrix}$$

is called the characteristic impedance matrix of the coupler. It states merely that, at the design frequency, Z_0 is the impedance presented to both the main and auxiliary waveguides, and that both waveguides are completely decoupled at the input and output ends of the coupler.

In this form the coupler is not ideally suited for cascading with other couplers of differing coupling factors since the image impedances are a function of the stub impedances and these vary with the coupling factor. Fig. 2(a) shows how the impedances may be rearranged to make the image impedances unity irrespective of

(a)

(b)

Fig. 2.—Impedances in a matched 2-stub coupler.

(a) Impedances expressed in terms of the power coupling factor
  c = Power coupling factor for the coupler.
(b) Impedances expressed in terms of θ.
  sin² θ = Power coupling factor of the coupler.

the coupling. It is now possible to cascade as many couplers of differing coupling factors as may be necessary, and each will be correctly terminated.

(2.2) The Voltage Transfer Matrix

For the coupler shown in Fig. 2(a) the voltage transfer matrix relating the output voltages to the input voltages for the 2-stub coupler is

$$\begin{bmatrix} V_1'' \\ V_2'' \end{bmatrix} = \begin{bmatrix} \sqrt{(1-c)} & -j\sqrt{c} \\ -j\sqrt{c} & \sqrt{(1-c)} \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \quad . \quad . \quad . \quad (4)$$

If the reference phase in the auxiliary guide is retarded by 90°, the j terms in the matrix become real and it reduces to the simple form

$$\begin{bmatrix} \sqrt{(1-c)} & \sqrt{c} \\ -\sqrt{c} & \sqrt{(1-c)} \end{bmatrix}$$

Its simplicity becomes more apparent if we write

$$c = \sin^2 \theta \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

i.e. $\sin \theta$ is the voltage coupling factor for the coupler.

The transfer matrix then beomes

$$\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

If we define, in the usual way, the unit matrix

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and the J matrix

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

the voltage transfer matrix may be written as

$$U \cos \theta + J \sin \theta$$

Since the J matrix has the following properties:

$$J^2 = -U, \quad J^3 = -J, \quad J^4 = U, \text{ etc.}$$

the voltage transfer matrix has an algebra identical with that of the ordinary complex number

$$\cos \theta + j \sin \theta$$

In particular, it may be written in the form

$$\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} = \exp(J\theta)$$

making use of de Moivre's theorem.

Fig. 2(b) shows, in schematic form, the 2-stub coupler with the waveguide characteristic impedances expressed in terms of θ.

(3) CASCADED 2-ELEMENT COUPLERS

The voltage transfer matrix has properties analogous to those of the propagation coefficient in the theory of lossless iterative 2-terminal-pair networks.

Thus if two 2-stub couplers having voltage coupling factors, $\sin \theta_1$ and $\sin \theta_2$, are cascaded, the overall voltage transfer matrix for the resultant 4-slot coupler is obtained by multiplying together the individual transfer matrices using de Moivre's theorem.

Thus $\quad \exp(J\theta 1) \times \exp(J\theta_2) = \exp J(\theta_1 + \theta_2)$

$$= \begin{bmatrix} \cos(\theta_1 + \theta_2) & \sin(\theta_1 + \theta_2) \\ -\sin(\theta_1 + \theta_2) & \cos(\theta_1 + \theta_2) \end{bmatrix}$$

The overall power coupling factor for the 4-slot coupler is thus

$$\sin^2(\theta_1 + \theta_2)$$

Miller has given expressions for the voltages in the main and auxiliary waveguides [eqns. (39) and (40) on p. 694 of Reference 3] which are merely an expanded version of de Moivre's theorem.

(3.1) A Semi-Infinite Chain of Identical 2-Stub Couplers

Consider a semi-infinite chain of 2-stub couplers each having a voltage coupling factor $\sin\theta$. The voltage transfer matrix for the whole structure is

$$\begin{bmatrix} \cos n\theta & \sin n\theta \\ -\sin n\theta & \cos n\theta \end{bmatrix}$$

It is apparent that, if we start at the generator end of the structure with energy in the main waveguide only and proceed down the chain, the amplitudes of the voltages in the main and auxiliary waveguides will vary cosinsuoidally and sinusoidally, respectively. This is illustrated in Fig. 3.

Fig. 3.—Variation of the voltage in the main and auxiliary waveguides in passing down a chain of cascaded 2-stub couplers.

If, in the limit, the coupling factor of each 2-stub element is made indefinitely small, any coupling factor can be obtained by terminating the structure after an appropriate number of sections. Alternatively, given a coupling factor and a required number of sections, the coupling factor per section may readily be calculated.

(3.2) Binomial Slot Couplers

Lippmann[1] and others have dealt with the first-order theory of broad-band directional couplers in which the voltage coupling factors of the elements are so proportioned that they form a binomial sequence when normalized with respect to the coupling factors of the smallest elements. The basis of the theory may be quoted from Lippmann[1] as follows:

> Let us consider two transmission lines joined by N arbitrary, equally spaced coupling elements and let us call the backward, or undesired voltage wave coupled from one of the lines into the other V. Then, on the basis of first-order theory, it may be shown that the necessary and sufficient condition that V and its first $(N-2)$ derivatives with respect to the frequency be zero at the design frequency is that the N elements form a binomial coupler. The nature of the coupling elements is completely arbitrary; they may be holes, slots, T-junctions, probes or combinations of these, but the elements must couple proportionally in both directions.

Although this theory is satisfactory for couplers having weak overall coupling (say 10 dB or more) it may be shown that there are serious objections to its validity in the case of strong coupling.

Nevertheless 3 dB couplers have been designed on this princip and have performed satisfactorily.[4]

In the following Section an example is given of the design such a coupler.

(3.2.1) The Design of a Binomial-Slot 3 dB Coupler.

Consider a cascaded pair of identical 2-stub couplers. It clear from the fact that the match and directivity of these couple are perfect at the design frequency that the performance of th resultant 4-slot coupler at the design frequency is independent the distance between the individual couplers. At other fr quencies, however, the directivities are no longer perfect ar there will be two backward waves set up in the auxiliary wav guide. If it is assumed that the coupling is weak, the first-ord theory can be applied, and use can be made of the fact that thes backward waves will destroy one another if the distance betwee the couplers is an odd number of quarter wavelengths. Th least frequency-sensitive design is obviously obtained when th distance is one-quarter wavelength. The two middle slots the coincide, and they must be combined into a single slot havir twice the voltage coupling factor of each individual slot to give 3-slot binomial coupler (i.e. the stub impedances are as 1 : 2 : 1

The design of a binomial-slot 3 dB coupler for No. 16 wavegui is next considered. A 5-slot design is assumed in which the stu characteristic impedances are in the proportions 1 : 4 : 6 : 4 :

If the argument is extended for the 3-slot binomial coupler, is readily shown that the 5-slot coupler may be broken dow into a cascaded set of four 2-stub couplers with voltage couplin factors in the proportions 1 : 3 : 3 : 1.

Let the voltage coupling for the two end sections be

$$\sin\theta_1 = x$$

and for the inner sections

$$\sin\theta_2 = 3x$$

For the 3 dB coupler we have for the overall voltage transf matrix

$$\begin{bmatrix} \cos 2(\theta_1 + \theta_2) & \sin 2(\theta_1 + \theta_2) \\ -\sin 2(\theta_1 + \theta_2) & \cos 2(\theta_1 + \theta_2) \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\ -\dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \end{bmatrix}$$

i.e.

$$\theta_1 + \theta_2 = \frac{\pi}{8} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (6$$

If x is assumed to be sufficiently small, so that

$$\left. \begin{array}{l} \theta_1 \simeq x \\ \theta_2 \simeq 3x \end{array} \right\} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (7$$

an approximate solution for x may be found at once.

Substituting in eqn. (6), we have

$$4x = \frac{\pi}{8}$$

$$x = \frac{\pi}{32} = 0 \cdot 098$$

$$3x = 0 \cdot 294$$

However, an exact solution for x may be found without difficult From eqn. (6)

$$\theta_1 = \frac{\pi}{8} - \theta_2$$

ow $$3 \sin \theta_1 = \sin \theta_2$$

therefore $$3 \sin \left(\frac{\pi}{8} - \theta_2 \right) = \sin \theta_2$$

i.e. $$3 \sin \frac{\pi}{8} \cos \theta_2 - 3 \cos \frac{\pi}{8} \sin \theta_2 = \sin \theta_2$$

$$\tan \theta_2 = \frac{3 \sin \dfrac{\pi}{8}}{1 + 3 \cos \dfrac{\pi}{8}}$$

$$\theta_2 = 16° \ 56'$$

therefore $$\sin \theta_2 = 0 \cdot 291$$

$$\sin \theta_1 = 0 \cdot 097$$

the ideal configuration of impedances shown in Fig. 2(b) is sumed, the height b' of the stubs must be such as to give an effective characteristic impedance $\tan \theta_1$ for the stubs in the end ctions, and $\tan \theta_2$ for the stubs in the middle sections. The quired values of b' may be found from a curve in Harrison's port,[2] which gives the relationship between b' and the coupling r 2-stub couplers based on the assumption that the coupling is log Z^2 decibels, where Z is the effective characteristic imped- ce of the stubs after corrections have been made to allow for e effect of the T-junctions.

In this case the coupling factors for entering this curve are follows:

$$C_1 = 10 \log (\tan^2 \theta_1) = -20 \cdot 2 \, \text{dB}$$

$$C_2 = 10 \log (\tan^2 \theta_2) = -10 \cdot 3 \, \text{dB}$$

hich give the corresponding values of b':

$$b'_1 = 0 \cdot 040 \, \text{in}$$

$$b'_2 = 0 \cdot 130 \, \text{in}$$

he heights of the main and auxiliary guides between the stubs e found as follows:

$$b_1 = b \sec \theta_1 = 0 \cdot 400 \times 1 \cdot 005 = 0 \cdot 402 \, \text{in}$$

$$b_2 = b \sec \theta_2 = 0 \cdot 400 \times 1 \cdot 045 = 0 \cdot 418 \, \text{in}$$

he corrections to the quarter-wavelength distances between the ntre-lines of the stubs and between the main and auxiliary uides are readily found from the curves given by Harrison.

The complete dimensions of the 2-stub sections are shown in g. 4, based on a design frequency of 9 368 Mc/s ($\lambda_g = 4 \cdot 48$ cm). he 3 dB coupler may be assembled in three ways from these asic elements, namely

(a) Spacing the individual 2-stub couplers $3\lambda_g/4$ apart, centre-line cto centre-line,

(b) Spacing the two middle elements $\lambda_g/4$ apart and connecting the resultant 1 : 2 : 1 coupler to the two end-sections by waveguides $\lambda_g/2$ in length,

(c) Spacing all four 2-stub couplers $\lambda_g/4$ apart to give a 5-slot binomial coupler. The separation of the main and auxiliary guides then becomes a matter of compromise.

The author has not had an opportunity of checking the relative erformance of these three types of coupler, but the dimensions the 5-slot 3 dB coupler whose performance is described in eference 4 are reproduced in Fig. 5. No attempt was made in is design to apply corrections to the heights of the main and uxiliary waveguides, but in spite of this, the experimentally etermined slot widths bear a striking resemblance to those btained by directly cascading the 2-stub sections derived above.
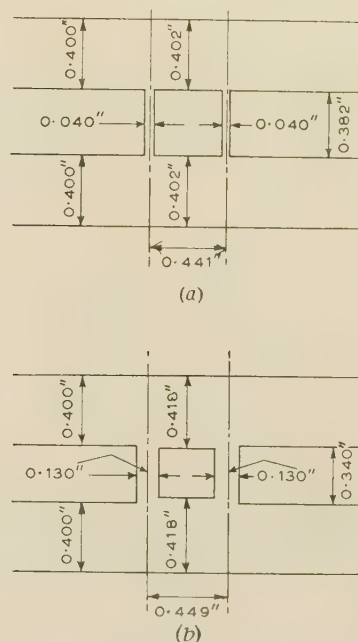
Fig. 4.—Dimensions of the 2-stub sections.
(a) End-sections. Coupling factor = 20·2 dB.
(b) Middle-sections. Coupling factor = 10·3 dB.
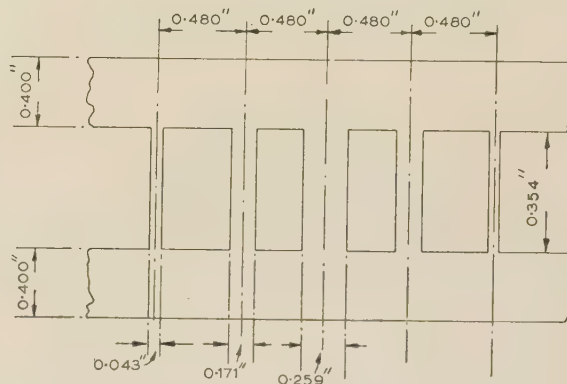
Fig. 5.—Dimensions of 5-slot binomial hybrid.

Thus we have 0·043 in, 0·171 in and 0·259 in, respectively, for the slots in the experimental coupler and 0·040 in, 0·170 in, and 0·260 in, respectively, for the coupler designed by the method described above.

(3.2.2) The Effect of Machining Tolerances on the Slot Widths.

The design procedure outlined in the preceding Sections may be used to find the effect of machining tolerances on the slot widths on the overall coupling of a multi-slot coupler.

For this purpose we may use the approximation

$$\theta \simeq \frac{b'}{b}$$

Taking, as an example, the 5-slot 3 dB coupler described in the preceding Section it may be assumed that the tolerances on the slot widths for the 2-stub elements are $\pm 0 \cdot 001$ in. The maxi-

mum and minimum possible voltage coupling factors for the 2-stub elements are given by

$$\frac{\theta_{max}}{\theta_{min}} \simeq \frac{b' \pm 0 \cdot 001 \text{ in}}{b} = \frac{b'}{b} \pm 0 \cdot 002 \, 5$$

θ_{max} refers to the plus and θ_{min} to the minus sign.

Assuming the slot widths are all on 'top limit', the voltage transfer matrix becomes

$$\begin{bmatrix} \cos\left(\frac{\pi}{4} + 0 \cdot 01\right) & \sin\left(\frac{\pi}{4} + 0 \cdot 01\right) \\ -\sin\left(\frac{\pi}{4} + 0 \cdot 01\right) & \cos\left(\frac{\pi}{4} + 0 \cdot 01\right) \end{bmatrix} = \begin{bmatrix} 0 \cdot 700 \ 0 \cdot 714 \\ -0 \cdot 714 \ 0 \cdot 700 \end{bmatrix}$$

The error in output balance is then

$$10 \log_{10}\left[\frac{0 \cdot 714}{0 \cdot 700}\right]^2 = 0 \cdot 18 \text{ dB}$$

In this case the overall tolerances on the three middle slots of the coupler, which each comprise a pair of slots from adjacent 2-stub elements, will be $\pm 0 \cdot 002$ in. If this tolerance is reduced to $\pm 0 \cdot 001$ in and the tolerance on the end slots is made $0 \cdot 000 \, 5$ in, the greatest error in output balance from this cause will be less than $0 \cdot 1$ dB.

(4) CONCLUSIONS

It has been shown how de Moivre's theorem may be used in conjunction with correction curves for end effects given by Harrison[2] to simplify the design of multi-element directional couplers of the branched-guide type.

The method also makes possible a simple assessment of th effect of manufacturing tolerances on the slot widths on th overall coupling factors.

As an example the design of a 5-slot 3 dB coupler was discussed the dimensions of the slots derived by this method agree well with those adjusted experimentally to give 3 dB coupling.

(5) ACKNOWLEDGMENTS

(6) REFERENCES

(1) LIPPMANN, B. A.: 'Theory of Directional Couplers' (Massa chusetts Institute of Technology Radiation Laborator Report No. 860).
(2) HARRISON, R. J.: 'Design Considerations for Directiona Couplers' (Massachusetts Institute of Technology Radia tion Laboratory Report No. 724).
(3) MILLER, S. E.: 'Coupled Wave Theory and Waveguid Applications', *Bell System Technical Journal*, 1954, **33** p. 661.
(4) LOMER, P. D., and CROMPTON, J. W.: 'A New Form o Hybrid Junction for Microwave Frequencies', *Proceeding I.E.E.*, Paper No. 2356 R (**104** B, p. 261).

SOME CHARACTERISTICS OF DOUBLE-EXPONENTIAL PULSE-SHAPING NETWORKS IN HIGH-VOLTAGE IMPULSE GENERATORS

By G. ELLESWORTH, B.Sc.(Eng.), Associate Member.

SUMMARY

The influence of the magnitude of network parameters on the waveform generated by the simplest types of circuit used in high-voltage impulse testing has been discussed at some length in the literature. These treatises are based, however, on simplifying assumptions not absolutely necessary, and no attempt appears to have been made to develop a unified approach to this problem. Moreover, some important circuits of slightly greater complexity have received particularly scanty attention, ostensibly on the grounds of mathematical complication.

The present paper introduces a unified treatment, without approximations, and with application to the more complex as well as to the simplest circuits. It indicates how one of the former may be employed to achieve higher efficiencies, coupled with increased economy of parts, particularly where resistive potential dividers are used. Some light is also thrown on the discrepancy between the nominal and actual rise times of double-exponential impulse waves.

LIST OF SYMBOLS

R_s, R_1, R_2 = Series resistance elements in a pulse-shaping circuit.

R_p, R_3, R_4 = Parallel resistance elements in a pulse-shaping circuit.

$r_s, r_p, r_1, r_2, r_3, r_4$ = Normalized values of network resistances.

C_s = Equivalent output capacitance of the impulse generator.

C_p = Equivalent 'load-circuit' capacitance.

i_1, i_2, i_3 = Instantaneous values of currents flowing in different parts of the pulse-shaping circuit.

I_{10}, I_{20}, I_{30} = Initial values of i_1, i_2 and i_3.

t = Time variable.

T_1 = Actual rise time of double-exponential impulse wave.

T_{1A} = Nominal rise time of double-exponential impulse wave calculated in accordance with A.S.A. 68.1 : 1953.

T_{1B} = Nominal rise time of double-exponential impulse wave calculated in accordance with B.S. 923 : 1940.

T_2 = Time to half-value of double-exponential impulse wave.

T_3, T_4, T_5 = Times in which double-exponential impulse wave rises to 10%, 90% and 30% of its peak value, respectively.

τ = Normalized time variable.

$K = T_2/T_1$.

$K_A = T_2/T_{1A}$.

$K_B = T_2/T_{1B}$.

$\rho_A = T_1/T_{1A}$.

$\rho_B = T_1/T_{1B}$.

V_0 = Initial voltage to which C_s is charged.

v = Impulse voltage across test object as a function of time.

v_{max} = Peak value of impulse voltage across test object.

η_v = Voltage efficiency.

λ = Waveform-dependent efficiency parameter.

(1) INTRODUCTION

The two simplest forms of double-exponential pulse-shaping network are shown in Figs. 1 and 2. They will be recognized as limiting cases of the circuit illustrated in Fig. 3; circuit A can
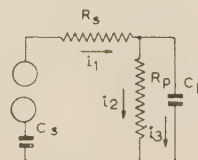
Fig. 1.—Circuit A: Conventional non-inductive equivalent circuit of an h.v. impulse generator with two energy-storage regions.
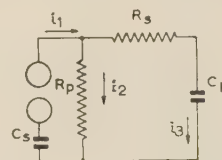
Fig. 2.—Circuit B: Alternative form of simplified equivalent circuit of an h.v. impulse generator with two energy-storage regions.
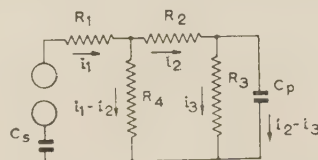
Fig. 3.—Circuit C: Generalized non-inductive equivalent circuit of an h.v. impulse generator with two energy-storage regions.

be obtained from circuit C by making $R_4 = \infty$ and $R_2 = 0$, while circuit B results from making $R_1 = 0$ and $R_3 = \infty$. In all these circuits, C_s represents the effective capacitance of the impulse generator, while the other circuit elements are used in waveform shaping. They incorporate such resistive and capacitive stray components as arise within the generator, the measuring circuit and the test object. Circuit inductance should be kept to a minimum by careful design and proper disposition of all parts of the network; the influence of residual inductance can then be taken into consideration in the manner described in the literature.[1,2,3,4]* The effective impulse voltage wave is in each case developed across C_p.

* The testing of highly inductive circuits, such as transformers, merits special treatment, South Australia.

The influence of the magnitude of network parameters on the impulse waveform has been discussed at great length in the literature.[1-6] These treatises are based, however, on certain simplifying assumptions, and no attempt appears to have been made to develop a unified approach facilitating effective design and optimum performance of high-voltage pulse-shaping networks. Moreover, the circuit in Fig. 3 receives particularly scanty attention in these references, ostensibly on the grounds of mathematical complication. Its performance has, in consequence, not been analysed sufficiently to illustrate the manner in which it may produce optimum results. Finally, with one exception,[6] none of the references available deals with the discrepancy between the theoretical rise time and that calculated in accordance with British,[7] American[8] or I.E.C.[9] Standards. It is the purpose of the paper to offer a unified treatment which includes all three circuits and to provide simple means whereby the most satisfactory operating conditions can be determined in both the technical and economic senses.

(2) NORMALIZED WAVEFORM EQUATIONS

The curve OAB in Fig. 4 is typical of double-exponential impulse waves, except for the initial portion which is here shown concave upwards for convenience of drawing. Such waves are specified by three parameters: the rise time, T_1, the time to half-value, T_2, and the peak value of the voltage, v_{max}. For convenience of measurement B.S. 923:1940 defines the nominal value of T_1 as $T_{1B} = 1 \cdot 25(T_4 - T_3)$, while the American Standard A.S.A. C68.1:1953 defines it as $T_{1A} = 2(T_4 - T_5)$. In each case T_3, T_4 and T_5 are the times shown in Fig. 4. Any network with
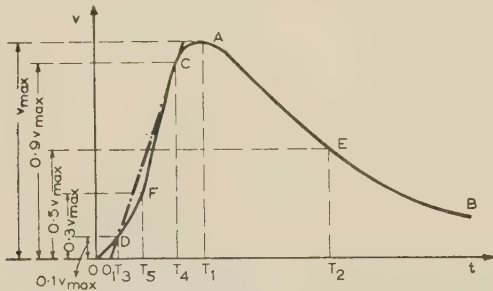
Fig. 4.—Double-exponential impulse-voltage waveform.

two energy storage regions will lend itself to the generation of such waves. The basic equations of double-exponential impulse waves are derived in normalized form in Section 7.1 where it is shown that

$$v = A(\varepsilon^{-\alpha\tau} - \varepsilon^{-\beta\tau}) \quad . \quad . \quad . \quad . \quad (1)$$

α and β being constants related by the transcendental equations

$$\beta = \alpha\varepsilon^{\beta - \alpha} \quad . \quad . \quad . \quad . \quad (2)$$

and

$$\tfrac{1}{2}(\varepsilon^{-\alpha} - \varepsilon^{-\beta}) = \varepsilon^{-K\alpha} - \varepsilon^{-K\beta} \quad . \quad . \quad . \quad (3)$$

K is the ratio of time to half-value to actual rise time, i.e.

$$K = T_2/T_1 \quad . \quad . \quad . \quad . \quad . \quad (4)$$

while τ is the *normalized time* in which the actual rise time T_1 is used as the unit of time (instead of the microsecond).

Thus

$$\tau = t/T_1 \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and

$$A = v_{max}/(\varepsilon^{-\alpha} - \varepsilon^{-\beta}) \quad . \quad . \quad . \quad . \quad (6)$$

since $v = v_{max}$ when $\tau = 1$.

The solution, in closed form, of eqns. (2) and (3) is imprac-

ticable, but iterative solutions for β and K are easily obtained for any given value of α. The results of these computations are presented in Figs. 5 and 6.
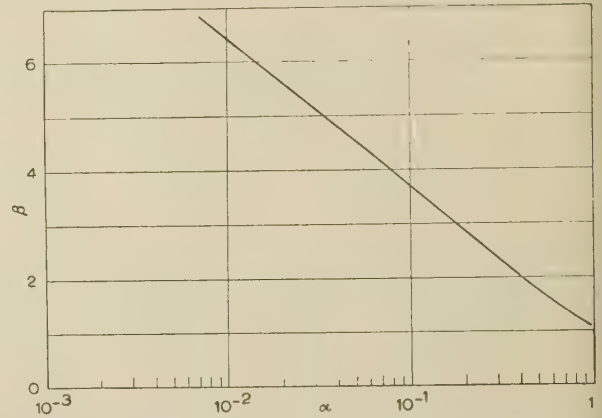
Fig. 5.—Relationship between the waveform parameters α and β.
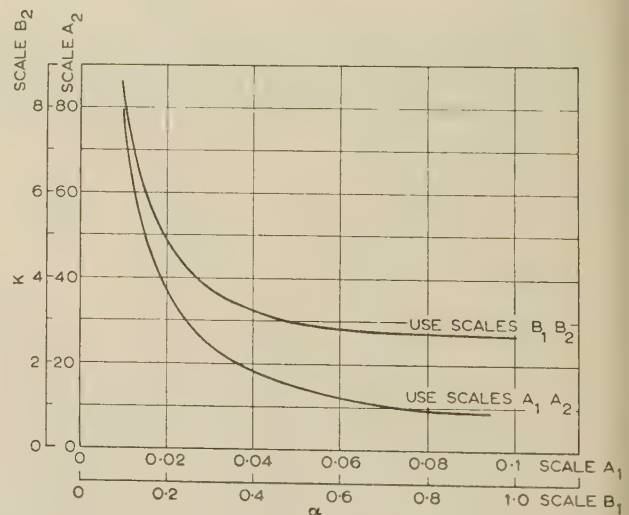
Fig. 6.—Relationship between the waveform parameters K and α.

The importance of these results lies in the observation that both α and β are determined, and with them the waveform, by K alone.

It may be emphasized again that these results will apply to all circuits possessing two energy-storage regions. In particular they apply to all three circuits in Figs. 1–3.

(3.1) Characteristics of Circuit A

Once the secondary parameters T_1 and K, and therefore α and β, are known, it is easy to determine the circuit constants in Fig. 1. For it follows from Section 7.1 that

$$\frac{\alpha + \beta}{T_1} = \frac{1}{R_p C_p} + \frac{1}{R_s C_p} + \frac{1}{R_s C_s} \quad . \quad . \quad (7)$$

and

$$\alpha\beta/T_1^2 = 1/R_s R_p C_s C_p \quad . \quad . \quad . \quad . \quad (8)$$

Upon eliminating R_s from these two equations a quadratic equation is obtained in R_p, the solution of which may be written in the form

$$r_p = \frac{R_p C_s}{T_1} = \frac{\alpha + \beta \pm \sqrt{[(\beta - \alpha)^2 - 4\alpha\beta N]}}{2\alpha\beta(N + 1)/N} \quad . \quad (9)$$

where

$$N = C_s/C_p \quad . \quad . \quad . \quad . \quad . \quad (10)$$

and r_p may be regarded as the 'normalized' value of R_p. Eqns. (8) and (9) may then be used to calculate R_s (or its normalized value r_s) in the form

$$r_s = \frac{R_s C_s}{T_1} = \frac{N/\alpha\beta}{R_p C_s/T_1} \quad . \quad . \quad . \quad (11)$$

We have already seen that specification of the wave by means of the ratio T_1/T_2 determines K, and with it α and β; one may now note that C_s and C_p, or C_s and N, must also be known before R_p and R_s can be calculated. Of these, the value of C_s is usually established from considerations involving the energy content of the impulse wave; C_s is in any case determined once and for all in a given generator. Thus for a known value of K, only N remains as an independent parameter. Apart from the desirability of using resistances and capacitances of convenient magnitude, some further criterion of circuit performance appears to be necessary so that one may find the best combination of primary circuit constants. The voltage efficiency, η_v, defined by the relationship

$$\eta_v = v_{max}/V_0 \quad . \quad . \quad . \quad . \quad . \quad (12)$$

may be used for this purpose, V_0 being the initial voltage across C_s. Its value is derived in Section 7.3, where it is shown that, for circuit A,

$$\eta_v = \lambda R_p C_s/T_1 \quad . \quad . \quad . \quad . \quad (13)$$

with

$$\lambda = \alpha\beta(\varepsilon^{-\alpha} - \varepsilon^{-\beta})/(\beta - \alpha) \quad . \quad . \quad (14)$$

Here λ is an efficiency parameter which is a function of the waveform but is independent of the circuit constants.
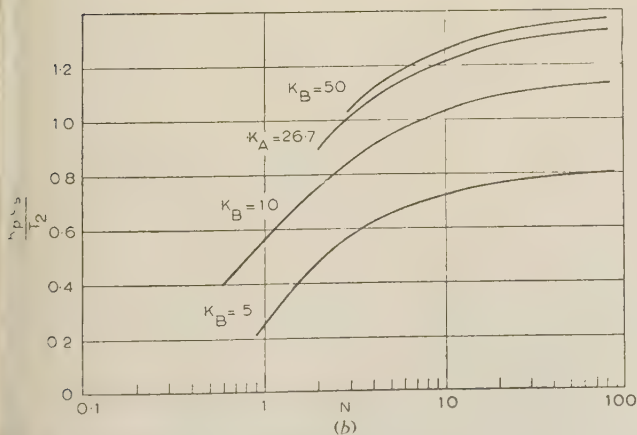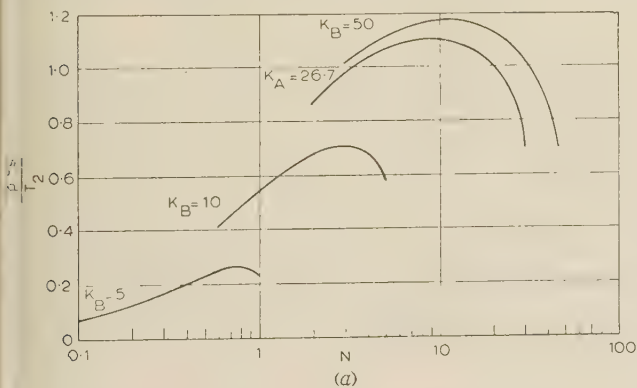
Fig. 7.—Variation of parallel resistance, R_p (normalized), with capacitance ratio, N, for four standard waveforms, using (a) circuit A; (b) circuit B.

From this it is clear that for any given waveform (i.e. $K = $ constant) η_v varies with N in the same way as R_p, and that for maximum voltage efficiency R_p should be made as large as possible. These conclusions are borne out by Figs. 7(a), 8(a) and 9(a), in the first two of which $R_p C_s/T_2$ and $R_s C_s/T_2$ are
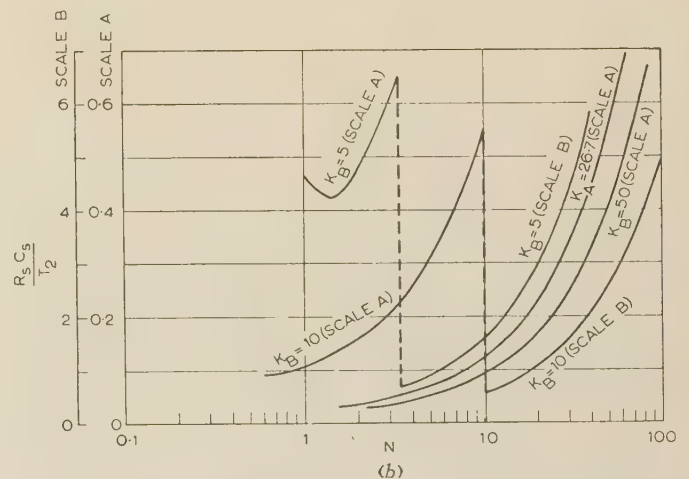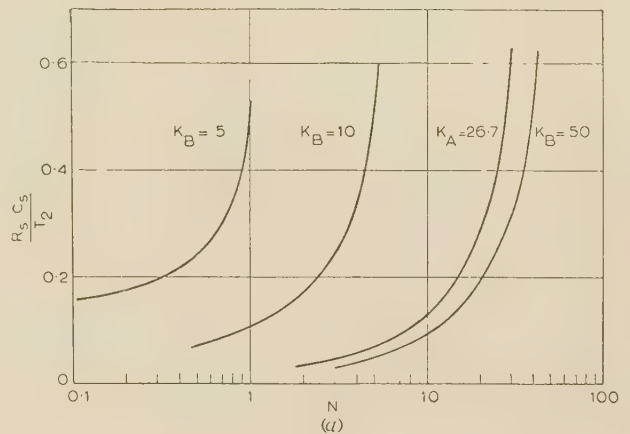
Fig. 8.—Variation of series resistance, R_s (normalized), with capacitance ratio, N, for four standard waveforms, using (a) circuit A; (b) circuit B.

plotted instead of $R_p C_s/T_1$ and $R_s C_s/T_1$ as the former lead to more convenient scales.

(3.2) Relation between Actual and Nominal Rise Times of Impulse Waves

In selecting the values of K for Figs. 7(a), 8(a) and 9(a) it was considered desirable to make allowance for the discrepancy between the actual and nominal rise times of impulse waves; the latter may be calculated in accordance with British, I.E.C., or American Standards,[7,8,9] These graphs were therefore plotted for the appropriate values of K_A and K_B (instead of K, as previously defined) where

$$K_A = \frac{T_2}{T_{1A}} = \frac{T_2}{2(T_4 - T_5)} \quad . \quad . \quad . \quad (15)$$

and

$$K_B = \frac{T_2}{T_{1B}} = \frac{T_2}{1 \cdot 25(T_4 - T_3)} \quad . \quad . \quad (16)$$

To do this it is necessary to establish the relationship between K and K_B or K_A respectively. For example, by putting $v = 0 \cdot 1\, v_{max}$
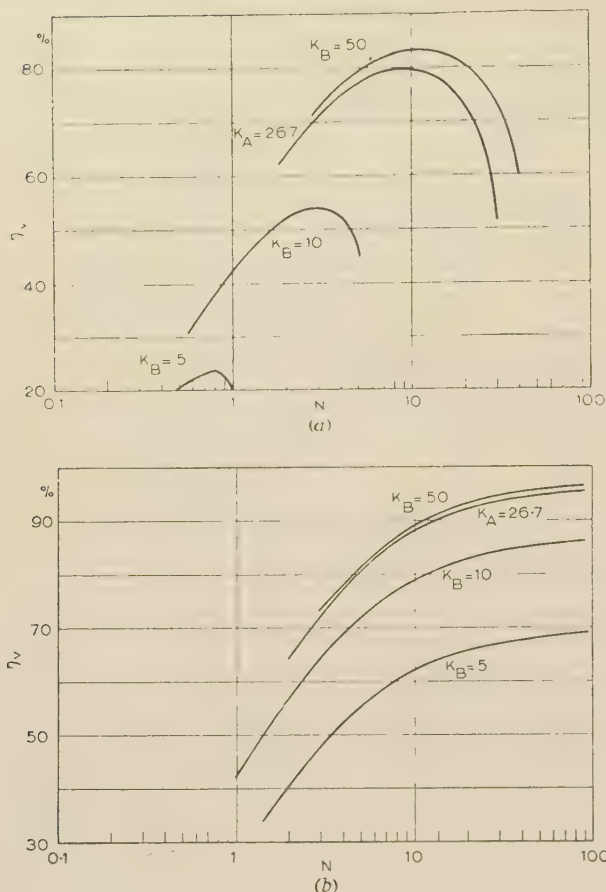
Fig. 9.—Variation of voltage efficiency, η_v, with capacitance ratio, N, for four standard waveforms, using (a) circuit A; (b) circuit B.

at $\tau = \tau_3$ and $v = 0.9v_{max}$ at $\tau = \tau_4$ in eqn. (31) (see Section 7.1) we obtain the transcendental equations

$$\varepsilon^{-\alpha\tau_3} - \varepsilon^{-\beta\tau_3} = 0.1(\varepsilon^{-\alpha} - \varepsilon^{-\beta}) \quad . \quad . \quad . \quad (17)$$

and $\qquad \varepsilon^{-\alpha\tau_4} - \varepsilon^{-\beta\tau_4} = 0.9(\varepsilon^{-\alpha} - \varepsilon^{-\beta}) \quad . \quad . \quad . \quad (18)$

From these τ_3 and τ_4 may be calculated by an iterative process for any value of K (and therefore α and β). As the actual rise time is always unity in our normalized system, the ratio of actual to nominal rise time, based on British[7] or I.E.C.[9] Standards, is

$$\rho_B = \frac{T_1}{T_{1B}} = \frac{1}{1.25(\tau_4 - \tau_3)} \quad . \quad . \quad . \quad (19)$$

Fig. 10 shows the variation of ρ_B with α.* The graphs in Figs. 6

* Similar data have been published by Angelini,[6] but there is considerable divergence between his results and those presented in this paper.
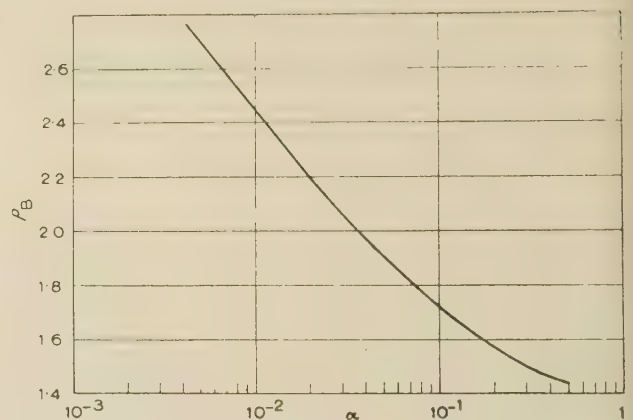
Fig. 10.—Ratio ρ_B of the actual rise time to the rise time calculated in accordance with B.S. 923:1940, as a function of the waveform parameter α.

and 10 may then be used to find α for any given K_B, since from eqns. (4), (16) and (19),

$$K_B = \rho_B K \quad . \quad . \quad . \quad . \quad . \quad (20)$$

A similar procedure may be used to find the variation of K with α; Table 1 summarizes the relevant data for the various standard waveforms.

The marked difference between T_1, T_{1B} and T_{1A}, amply demonstrated by Table 1, suggests the desirability of modifying the multiplying factors used in the present method of calculating the nominal rise times of standard impulse voltage waves, thus making them agree more closely with the actual rise times.

(3.3) Characteristics of Circuit B

On carrying out similar calculations in connection with circuit B, R_p is again found as the solution of a quadratic equation. Using the same notation as before, its value is now given by

$$r_p = \frac{R_p C_s}{T_1} = \frac{\alpha + \beta \pm \sqrt{[(\beta - \alpha)^2 - 4\alpha\beta/N]}}{2\alpha\beta(N + 1)/N} \quad . \quad (21)$$

which may be compared with eqn. (9). Eqns. (10) and (11) remain valid for this case without any modification at all. Likewise, it is shown in Section 7.3 that eqns. (13) and (14) also apply to circuit B, the voltage efficiency being defined by eqn. (12)

Table 1

DATA FOR STANDARD WAVEFORMS

Waveform	$K = T_2/T_1$	$K_B = T_2/T_{1B}$	$K_A = T_2/T_{1A}$	T_1	T_{1B}	T_{1A}	T_2	α	β
				µs	µs	µs	µs		
1/50	24·3	50	—	2·05	1	—	50	0·030	5·1817
$1\frac{1}{2} \times 40$	19	—	$26\frac{2}{3}$	2·09	—	$1\frac{1}{2}$	40	0·0396	4·8468
1/10	6·16	10	—	1·63	1	—	10	0·143	3·2738
1/5	3·41	5	—	1·47	1	—	5	0·365	2·1281

as previously. The families of curves in Figs. 7(b), 8(b) and 9(b) may therefore be drawn to illustrate the performance of this circuit.

(3.4) Conclusions concerning Circuits A and B

Comparison of corresponding graphs relating to circuits A and B leads to the following deductions:

(a) Circuit B can be used with a wider range of capacitance
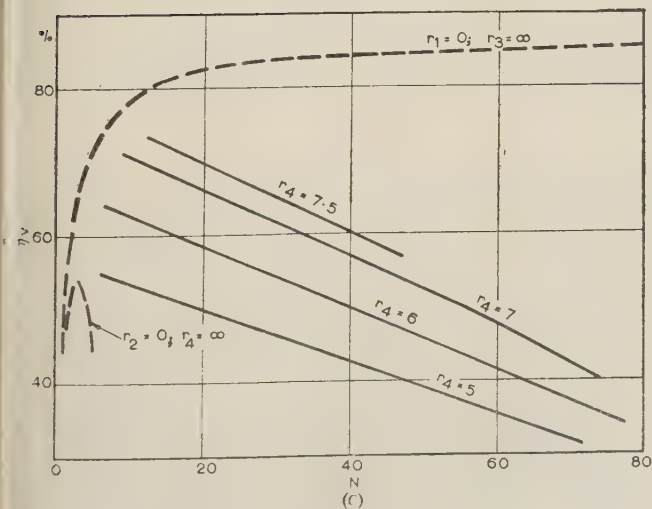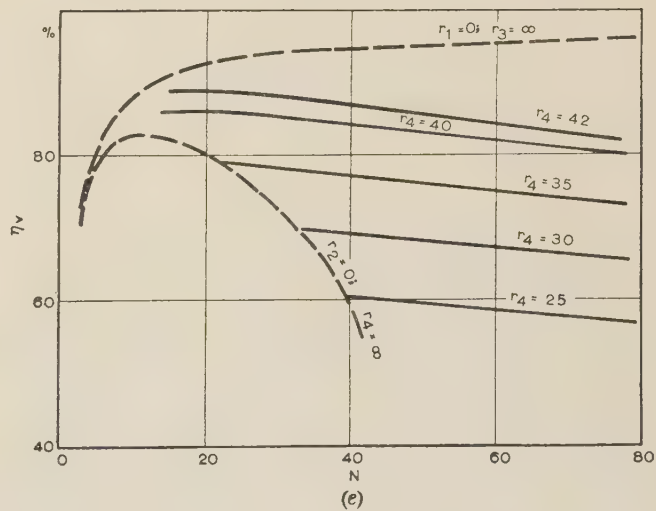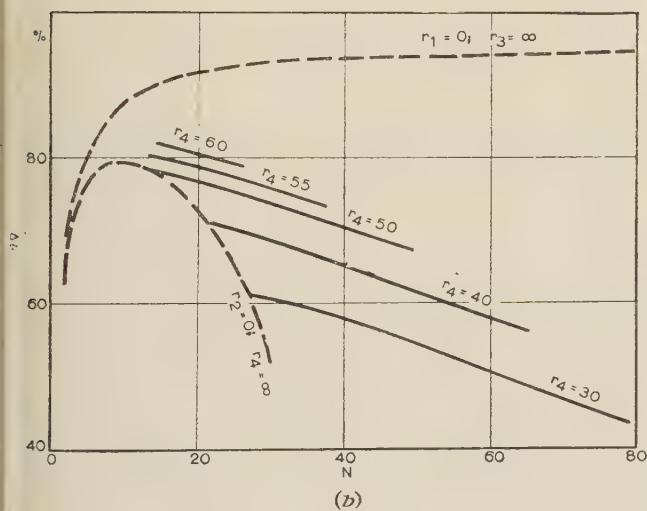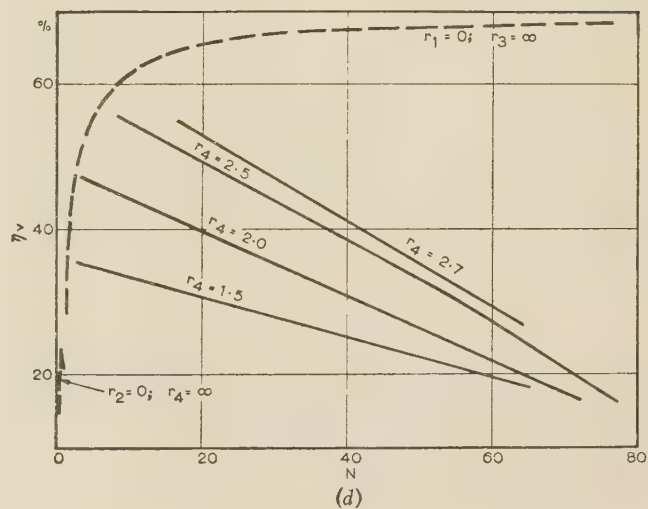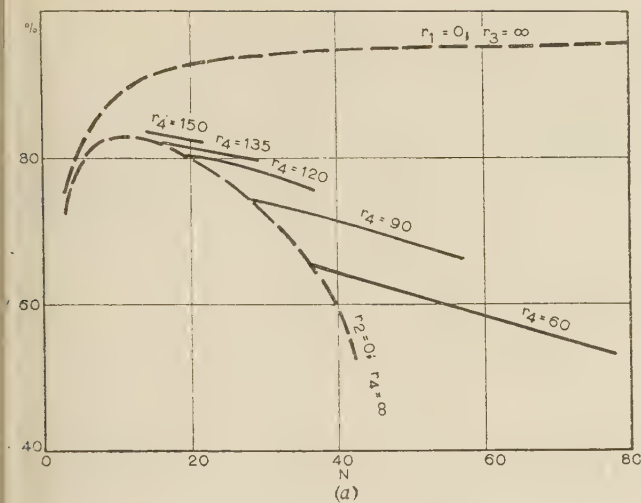
Fig. 11.—Variation of voltage efficiency, η_v, in circuit C with capacitance ratio, N, for different values of r_4.

(a) Standard   1/50 impulse wave ($K_B = 50$), $C_s = 5\,000\,\text{pF}$, $R_3 \simeq 15\,000\,\Omega$ ($r_3 = 36$).
(b) Standard 1½/40 impulse wave ($K_A = 26\cdot7$), $C_s = 5\,000\,\text{pF}$, $R_3 \simeq 15\,000\,\Omega$ ($r_3 = 35$).
(c) Standard   1/10 impulse wave ($K_B = 10$), $C_s = 5\,000\,\text{pF}$, $R_3 \simeq 15\,000\,\Omega$ ($r_3 = 45$).
(d) Standard   1/5   impulse wave ($K_B = 5$), $C_s = 5\,000\,\text{pF}$, $R_3 \simeq 15\,000\,\Omega$ ($r_3 = 50$).
(e) Standard   1/50 impulse wave ($K_B = 50$), $C_s = 5\,000\,\text{pF}$, $R_3 \simeq 50\,000\,\Omega$ ($r_3 = 120$).

ratios, N, than circuit A. This means that circuit B may give satisfactory performance with low values of C_p, where circuit A will either fail completely or work at a very low efficiency.

(b) The voltage efficiency of circuit B is higher than that of circuit A for all practical values of N (actually for $N > 1$). As against these advantages of circuit B, it must be remembered that its adoption will preclude the use of resistive potential dividers, which have certain advantages (as well as drawbacks) in the oscillographic measurement of impulse voltage waves. (The use of resistive voltage dividers is explicitly recommended by the current American Standard specifications.[8]) Thus the ability of circuit C to combine, in considerable measure, the advantages previously associated with circuit B and the facility of employing resistive voltage dividers for measuring purposes makes its use particularly attractive.

(4) CHARACTERISTICS OF CIRCUIT C

It is shown in Section 7.2 that the ratio, N, of the 'generator' and 'load' capacitances, and the 'generator resistance', R_1, are connected by the simultaneous equations

$$N = \alpha\beta\left[\frac{(r_2 + r_4)r_3}{r_2 + r_3 + r_4}r_1 + \frac{r_2 r_3 r_4}{r_2 + r_3 + r_4}\right] . \quad (22)$$

$$r_1 = \frac{\alpha + \beta}{\alpha\beta} - \frac{(r_2 + r_3)r_4}{r_2 + r_3 + r_4} - \frac{1}{N}\frac{(r_2 + r_4)r_3}{r_2 + r_3 + r_4} \quad (23)$$

where α and β have been defined before, and r_1, r_2, r_3 and r_4 are the 'normalized' values of the resistive components in Fig. 3, given by

$$r_k = R_k C_s/T_1; k = 1, 2, 3 \text{ or } 4 \quad . \quad . \quad (24)$$

It is also shown in Section 7.3 that, using the notation of eqn. (24), the voltage efficiency may in this case be written in the form

$$\eta_v = \lambda\frac{r_3 r_4}{r_2 + r_3 + r_4} \quad . \quad . \quad . \quad . \quad (25)$$

where λ is still the waveform-dependent efficiency parameter whose value is given by eqn. (14).

The simultaneous eqns. (22) and (23) may be used to evaluate N and r_1 for a given waveform if r_2, r_3 and r_4 are known, while eqn. (25) will give the efficiency under the same conditions. This method has been used in constructing the graphs shown in Figs. 11 and 12. Figs. 11(a)–11(d) corresponds to $R_3 \simeq 15\,000$ ohms ($C_s = 5\,000\,\text{pF}$) representing the resistance of a resistive potential divider of conventional design. Fig. 11(e) was drawn for $R_3 \simeq 50\,000$ ohms ($C_s = 5\,000\,\text{pF}$) and may be compared with Fig. 11(a) to show the influence of r_3 on network performance.

The following comments and conclusions are based on a study of these graphs:

(a) From Figs. 11(a)–11(d) it is clear that replacing circuit A by circuit C removes the restriction on the range of values of N at which the former can function.

(b) With the particular values chosen for the generator capacitance and the resistance of the potential divider ($R_3 \simeq 15\,000$ ohms, $C_s = 5\,000\,\text{pF}$) the maximum efficiency obtainable with circuit C is only slightly above that given by circuit A for the British Standard 1/50 and American Standard $1\frac{1}{2} \times 40$ waves; but it is very significantly increased, and almost equal to the relatively high efficiency associated with circuit B, for 1/10 and 1/5 waves.

(c) A comparison of Figs. 11(a) and 11(e) indicates that higher efficiencies, approaching those of circuit B, may be attained with

the 1/50 (and similarly with the $1\frac{1}{2} \times 40$) wave if higher values of r_3 are used.*

(d) The choice of shunt resistance R_4 (or in its normalized form r_4) is very important in attaining the highest efficiency with any waveform. It is obvious from all the relevant figures that r_4 should be made as high as practicable.

(e) The nature of the limitations upon indiscriminate increase in the value of r_4 is seen in Fig. 12, from which it appears that
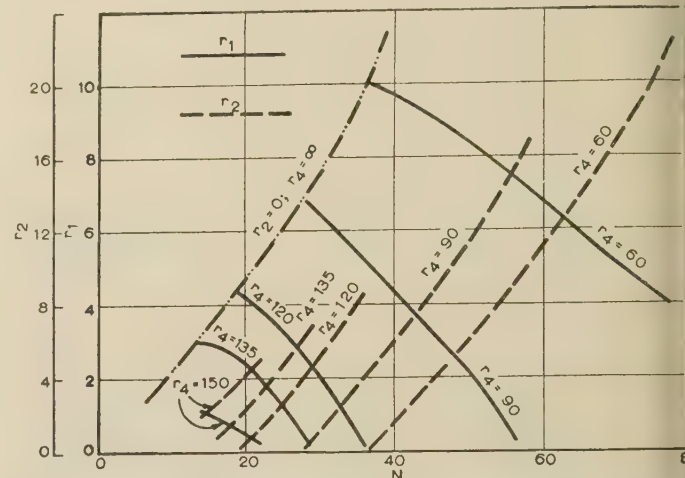
Fig. 12.—Variation of series resistances, r_1 and r_2 (normalized), in circuit C with capacitance ratio, N, for standard 1/50 impulse wave ($K_B = 50$), $C_s = 5\,000\,\text{pF}$, $R_3 \simeq 15\,000$ ohms ($r_3 = 36$) and different values of r_4.

for otherwise similar conditions, as r_4 is increased, r_1 decreases while r_2 becomes larger. For each waveform, and for a given value of r_3, there is a maximum value of r_4 which makes $r_1 = 0$; beyond this one would require $r_1 < 0$.†

(f) For all standard waves, capacitance ratios of the same magnitude, broadly $10 \leqslant N \leqslant 20$, appear to give optimum results when using circuit C. However, even if higher ratios are used, there is no appreciable deterioration of performance. It is possible, therefore, to obtain very reasonable results with a fixed set of shunt capacitors across the test object, or even relying on the stray capacitances across the potential divider and test piece alone.

Finally, it may be observed that, in their normalized form, Figs. 11 and 12 apply without modification to wavefronts of any duration, as long as K remains constant; e.g. Figs. 11(a) and 12 give the data for 0.1/5 as well as 1/50 waves. Moreover, for the same value of C_s and r_3, a reduction of T_1 will bring about a proportionate decrease in R_3, which is in accordance with the recommendations of the American Standards on this subject.[8]

(5) CONCLUSIONS

The foregoing results show that expressions can be derived without the use of simplifying assumptions that will give, on the one hand, the mathematical parameters α and β for any specified double-exponential impulse waveform, and on the other hand, appropriate values for the different circuit elements in each type of pulse-shaping network considered. By a simple process of normalization of the time and resistance scales, all equations, and the graphs plotted from them, have a general usefulness conferred upon them, instead of being applicable to one particular

* The effect of increasing R_3 on the accuracy of oscillographic measurement must not, however, be overlooked. It may be worth noting that the American Standards recommend that the potential-divider resistance should not exceed 20 000 ohms ever for rise times over 1 microsec duration.
† In practice it is found convenient, from the point of view of suppressing spurious oscillations within the generator, not to use extremely low values of r_1.

se only. In this normalized form, they show quite clearly the
xibility of the slightly more complex circuit in Fig. 3 and
licate the manner in which optimum technical performance
ay be obtained from it in conjunction with resistive potential
viders. At the same time, the possibility of operating the
pulse generator in a technically satisfactory manner with only
fixed set of shunt capacitors, or even entirely without such
acitors, is a matter of considerable economic importance.
As a side issue, it is noted that the method of calculating the
minal rise time of double-exponential impulse waves, as now
proved by the various national and international standards, is
mewhat unrealistic, as it leads to consistently large differences
tween the nominal and the true rise times of such waves.

(6) REFERENCES

1) EATON, J. R., and GEBELEIN, J. P.: 'Circuit Constants for
the Production of Impulse Test Waves', *General Electric
Review*, 1940, **43**, p. 322.

2) THOMASON, J. L.: 'Impulse Generator Circuit Formulas',
Electrical Engineering, 1934, **53**, p. 169.

3) THOMASON, J. L.: 'Impulse Generator Voltage Charts for
Selecting Circuit Constants', *ibid.*, 1937, **56**, p. 183.

4) CRAGGS, J. D., and MEEK, J. M.: 'High Voltage Laboratory
Technique' (Butterworths, 1954).

5) 'Surge Phenomena' (E.R.A. Report Ref. S/T35; 1941).

6) ANGELINI, A. M.: 'New Formulae for Practical Calculation
of Impulse Waves', C.I.G.R.É., Paris, 1950, Paper No. 335.

7) B.S. 923:1940.

8) A.S.A. C.68.1:1953.

9) I.E.C. 60:1938.

0) V.D.E. 0450:1939.

1) EDWARDS, F. S., HUSBANDS, A. S., and PERRY, F. R.: 'The
Development and Design of High-Voltage Impulse Gene-
rators', *Proceedings I.E.E.*, Paper No. 1089 M, January,
1951 (**98**, Part I, p. 155).

(7) APPENDICES

**(7.1) Normalization of the Waveform Equation and the Relation-
ship between the Parameters α and β**

The class of passive linear lumped-circuit networks possessing
two energy-storage elements, to which the three circuits in
Figs. 1–3 belong, have their response to a step-function deter-
mined by the differential equation

$$\ddot{v} + a\dot{v} + bv = 0 \quad . \quad . \quad . \quad . \quad (26)$$

here v is the voltage across C_p and the values of the coefficients
and b for each of the three circuits is given in Table 2.

The general solution of eqn. (26), assuming $v = 0$ when
$t = 0$, is

$$v = A\left[\varepsilon^{-\frac{1}{2}(a-P)t} - \varepsilon^{-\frac{1}{2}(a+P)t}\right] . \quad . \quad . \quad (27)$$

where

$$P = \sqrt{(a^2 - 4b)} \quad . \quad . \quad . \quad . \quad (28)$$

If α and β are defined by

$$\alpha = \tfrac{1}{2}(a - P)T_1 \quad . \quad . \quad . \quad . \quad (29)$$

and

$$\beta = \tfrac{1}{2}(a + P)T_1 \quad . \quad . \quad . \quad . \quad (30)$$

where, in the notation of Fig. 4, $dv/dt = 0$ and $v = v_{max}$ at
$t = T_1$, then

$$v = A(\varepsilon^{-\alpha\tau} - \varepsilon^{-\beta\tau}) = \frac{v_{max}}{\varepsilon^{-\alpha} - \varepsilon^{-\beta}}(\varepsilon^{-\alpha\tau} - \varepsilon^{-\beta\tau}) \quad (31)$$

and

$$\tau = t/T_1 . \quad . \quad . \quad . \quad . \quad (32)$$

Also, once again in the notation of Fig. 4,

$v = \tfrac{1}{2}v_{max}$ when $t = T_2 = KT_1$, i.e. when $\tau = K$; hence

$$\tfrac{1}{2}(\varepsilon^{-\alpha} - \varepsilon^{-\beta}) = \varepsilon^{-K\alpha} - \varepsilon^{-K\beta} \quad . \quad . \quad (33)$$

Lastly, since $dv/dt = 0$ when $\tau = 1$, we obtain from eqn. (31)

$$\beta = \alpha\varepsilon^{\beta-\alpha} \quad . \quad . \quad . \quad . \quad (34)$$

(7.2) Computation of the Primary Circuit Parameters

Taking the values of a and b for circuit C from Table 2 and
putting

$$N = C_s/C_p \quad . \quad . \quad . \quad . \quad (35)$$

we may write

$$N = bC_s^2R_3\frac{R_1R_2 + R_1R_4 + R_2R_4}{R_2 + R_3 + R_4} \quad . \quad . \quad (36)$$

and

$$R_1 = \frac{a}{bC_s} - \frac{(R_2 + R_3)R_4}{R_2 + R_3 + R_4} - \frac{1}{N}\frac{(R_2 + R_4)R_3}{R_2 + R_3 + R_4} \quad . \quad (37)$$

Then, using eqns. (29) and (30), we obtain

$$N = \alpha\beta\left(\frac{C_s}{T_1}\right)^2\left[\frac{(R_2 + R_4)R_3}{R_2 + R_3 + R_4}R_1 + \frac{R_2R_3R_4}{R_2 + R_3 + R_4}\right] \quad (38)$$

and

$$R_1 = \frac{\alpha + \beta}{\alpha\beta}\frac{T_1}{C_s} - \frac{(R_2 + R_3)R_4}{R_2 + R_3 + R_4} - \frac{1}{N}\frac{(R_2 + R_4)R_3}{R_2 + R_3 + R_4} \quad (39)$$

These equations may be expressed in normalized form by putting

$$r_\kappa = R_kC_s/T_1; k = 1, 2, 3 \text{ or } 4 \quad . \quad . \quad . \quad (40)$$

giving

$$N = \alpha\beta\frac{(r_2 + r_4)r_3}{r_2 + r_3 + r_4}r_1 + \frac{r_2r_3r_4}{r_2 + r_3 + r_4} \quad . \quad . \quad (41)$$

Table 2

VALUES OF a AND b FOR CIRCUITS A, B AND C

	a	b
Circuit A	$\dfrac{1}{R_pC_p} + \dfrac{1}{R_sC_p} + \dfrac{1}{R_sC_s}$	$\dfrac{1}{R_sR_pC_sC_p}$
Circuit B	$\dfrac{1}{R_pC_s} + \dfrac{1}{R_sC_p} + \dfrac{1}{R_sC_s}$	$\dfrac{1}{R_sR_pC_sC_p}$
Circuit C	$\dfrac{R_1(R_2 + R_3 + R_4) + (R_2 + R_3)R_4}{C_pR_3(R_1R_2 + R_1R_4 + R_2R_4)} + \dfrac{(R_2 + R_4)R_3}{C_sR_3(R_1R_2 + R_1R_4 + R_2R_4)}$	$\dfrac{R_2 + R_3 + R_4}{C_sC_pR_3(R_1R_2 + R_1R_4 + R_2R_4)}$

and
$$r_1 = \frac{\alpha + \beta}{\alpha\beta} - \frac{(r_2 + r_3)r_4}{r_2 + r_3 + r_4} - \frac{1}{N}\frac{(r_2 + r_4)r_3}{r_2 + r_3 + r_4} \quad . \quad (42)$$

This pair of simultaneous equations in N and r_1 may be solved for any given values of r_2, r_3 and r_4, α and β being constant for a particular impulse waveform.

(7.3) Computation of the Voltage Efficiency

To obtain the voltage efficiency, η_v, defined by eqn. (12), we may write for the initial values of the currents i_1, i_2 and i_3, using the notation of Fig. 3 throughout,

$$I_{10} = V_0 \Big/ \left(R_1 + \frac{R_2 R_4}{R_2 + R_4}\right) = V_0 \frac{R_2 + R_4}{R_1 R_2 + R_1 R_4 + R_2 R_4} \quad (43)$$

$$I_{20} = \frac{R_4 I_{10}}{R_2 + R_4} = V_0 \frac{R_4}{R_1 R_2 + R_1 R_4 + R_2 R_4} \quad . \quad . \quad (44)$$

and
$$I_{30} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

Also, by using eqns. (31) and (45),

$$I_{20} = C_p \left(\frac{dv}{dt}\right)_0 = \frac{C_p(\beta - \alpha)}{T_1(\varepsilon^{-\alpha} - \varepsilon^{-\beta})} v_{max} \quad . \quad . \quad (46)$$

From eqns. (44) and (46) one obtains, also using eqns. (29) and (30),

$$\eta_v = \frac{v_{max}}{V_0} = \frac{(\varepsilon^{-\alpha} - \varepsilon^{-\beta})T_1 R_4}{(\beta - \alpha)C_p(R_1 R_2 + R_1 R_4 + R_2 R_4)}$$

$$= \lambda\left(\frac{R_3 C_s}{T_1}\right)\frac{R_4}{R_2 + R_3 + R_4} \quad . \quad . \quad . \quad (47)$$

where
$$\lambda = \alpha\beta(\varepsilon^{-\alpha} - \varepsilon^{-\beta})/(\beta - \alpha) \quad . \quad . \quad . \quad (48)$$

Eqn. (47) may again be written in normalized form using eqn. (40); thus

$$\eta_v = \lambda\frac{r_3 r_4}{r_2 + r_3 + r_4} \quad . \quad . \quad . \quad . \quad (49)$$

For the special cases of circuits A and B, one must put in eqn. (49) $r_2 = 0$ and $r_4 = \infty$ and $r_1 = 0$ and $r_3 = \infty$, respectively. This gives for both circuits, using the notation of Figs. 1 and 2 respectively,

$$\eta_v = \lambda\frac{R_p C_s}{T_1} \quad . \quad . \quad . \quad . \quad (50)$$

DISTORTION DUE TO THE MISMATCH OF TRANSISTORS IN PUSH-PULL AUDIO-FREQUENCY AMPLIFIERS

By K. W. GURNETT and R. A. HILBOURNE, B.Sc.

SUMMARY

In a class-B push-pull amplifier having balanced transformers, even-order harmonic distortion can be caused by differences in the characteristics of the two transistors. By considering the voltage gain of the two halves of the amplifier it is possible to obtain general expressions from which distortion caused by any particular characteristic can be calculated.

The paper is primarily concerned with audio-frequency amplification. It is necessary to consider harmonics outside this range because of the possibility of their giving rise to intermodulation products in the a.f. range.

The parameters giving rise to distortion can be divided into the frequency-independent and frequency-dependent types.

With regard to the frequency-independent parameters, the only significant factors are the current gain factor, the variation of base-collector current gain factor with emitter current and the extrinsic base resistance. In the case of the frequency-dependent parameters, which, in general, only become important at the higher audio frequencies, the significant factors are the collector capacitance, the variation of current gain with frequency and the carrier-storage effect.

In order to obtain low distortion, some matching of transistor characteristics is necessary. Methods are described whereby this can be achieved by fairly simple measurements.

Most of the work described in the paper is based on a low-power junction-type transistor. However, the methods used are quite general, and therefore applicable to other types of transistors.

LIST OF PRINCIPAL SYMBOLS

A_i = Amplifier current gain.
α = Emitter-collector current gain factor.
$\bar{\alpha}$ = Mean value of α.
α_0 = Low-frequency value of α.
α_z = Value of α at zero emitter current.
α_{cb} = Base-collector current gain factor.
B = Harmonic factor in a half sine-wave.
β = Slope of α/I_e curve.
C_c = Collector-base junction capacitance.
e = Charge of an electron.
f_α = Cut-off frequency in the common-base arrangement.
$f_{\alpha cb}$ = Cut-off frequency in the common-emitter arrangement.
f_g = Cut-off frequency of an amplifier.
G = Voltage gain of an amplifier.
H = Ratio of the operating frequency to f_g.
I_e, I_b and I_c = Emitter, base and collector currents, respectively.
i_s = Signal current.
K = Ratio of operating frequency to f_α.
k = Boltzman's constant.
m = Ratio of the values of r_{b0}.
n = Ratio of the values of G.
r_{b0} = Extrinsic base resistance.

r_{b1} = Intrinsic base resistance.
r_e = Emitter resistance.
r_c = Collector resistance.
R_S = Source resistance.
R_L = Load resistance.
R_{in} = Input resistance.
R_E = Total resistance in the emitter circuit.
T = Temperature.
v_L = Signal voltage across R_L.
v_s = Source signal voltage.
V_c = Collector voltage.

(1) INTRODUCTION

In power-amplifier applications, using junction transistors, it is often desirable to employ class-B push-pull circuit arrangements. The main reasons are the rather limited power-handling capabilities of transistors and the higher power efficiency obtained. In such a circuit arrangement the factors giving rise to harmonic distortion can be divided into two categories, namely:

(*a*) Non-linearities in the transistors.
(*b*) Differences between the two halves of the amplifier.

Type (*a*) gives rise to odd-order harmonic distortion[1] whereas (*b*) gives rise to even-order harmonic distortion. Assuming that the rest of the circuit is exactly balanced, the even-harmonic distortion is due to differences in transistor parameters. With two exactly similar transistors, the even-order distortion is zero; such a pair of transistors are said to be 'matched'.

At present, some transistors are matched solely on the basis of the current gain factor; however, this is not really sufficient. A better approach using an oscilloscopic display has been used,[4] in which the performance of the two transistors was compared under actual operating conditions.

The paper deals with the transistor parameter variations which give rise to even-harmonic distortion. These parameters can be separated into two types. The first is concerned with the parameters which do not involve frequency; they produce the distortion obtained at low frequencies. The second type is concerned with the frequency-dependent parameters; in general, these become important at higher audio frequencies. This is particularly true when the transistor is used in the common-emitter arrangement, and for this reason, considerable attention is paid to this circuit configuration.

In the present state of transistor production there is a considerable spread in their characteristics, and a random selection of transistors can result in substantial even-order harmonic distortion. Thus some degree of matching of transistors is required. If the factors which give rise to distortion are known, simple methods of matching can be derived; the method described in the paper consists in matching a voltage gain factor measured under conditions which fairly closely simulate those obtained in a practical amplifier.

When transistor class-B amplifiers (common emitter) are

operated at high audio frequencies considerable distortion can arise owing to different phase shifts in the transistors and also owing to carrier storage effects. The latter occur at the cross-over region of the fundamental waveform and cannot be over-come by matching.

In the paper the terms 'match' and 'mismatch' are used solely to describe the differences between the two halves of the transistor amplifier and do not refer to impedance matching.

(2) LOW AUDIO-FREQUENCY MATCHING CONSIDERATIONS

The amplification of a sine-wave signal by a class-B push-pull amplifier is considered. The two transistors amplify alternate halves of the input signal with voltage gains G_1 and G_2, respec-tively; G is defined as the ratio of the voltage developed across the load to the open-circuit generator voltage. The waveform appearing across the common load may be written

$$v_L = G_1[A \sin \omega t - AB_2 \cos 2\omega t - AB_4 \cos 4\omega t - \ldots$$
$$AB_q \cos q\omega t] - G_2[A \sin (\omega t + \pi)$$
$$- AB_2 \cos 2(\omega t + \pi) \ldots - AB_q \cos q(\omega t + \pi)] \quad . \quad (1)$$

where the input signal is given by $2A \sin \omega t$ (open-circuit generator voltage), and B_2, B_4, etc., are factors representing the amplitude of the harmonics present in a half sine-wave.

This expression is correct for all forms of G_1 and G_2, whether constant or not, but it is very cumbersome to use if they are functions of the input signal. Therefore, they are assumed to be constants. This approximation ignores the effect of intermodula-tion products, but these are likely to be very small.

Thus

$$v_L = (G_1 + G_2) A \sin \omega t + (G_2 - G_1)AB_2 \cos 2\omega t$$
$$+ (G_2 - G_1)AB_4 \cos 4\omega t + \ldots (G_2 - G_1)AB_q \cos q\omega t \quad . \quad (2)$$

From eqn. (2) the amplitude of the fundamental frequency is given by $(G_2 + G_1)A$ and the amplitude of the qth harmonic by $(G_2 - G_1)AB_q$. Thus the percentage qth harmonic content is given by

$$\frac{G_2 - G_1}{G_2 + G_1}B_q \times 100 \quad . \quad . \quad . \quad . \quad (3)$$

If the ratio of the gain factors G_1/G_2 is n, the percentage qth harmonic content is given by

$$\frac{1 - n}{1 + n}B_q \times 100 \quad . \quad . \quad . \quad . \quad (4)$$

The values of B_q for a half sine-wave are known. Thus

$$B_2 = 0\cdot425$$
$$B_4 = 0\cdot089$$
$$B_6 = 0\cdot038$$
$$B_8 = 0\cdot02$$
$$B_{10} = 0\cdot013$$

Each B term is multiplied by the same factor $(1 - n)/(1 + n)$; this can therefore be determined by measuring only one of the harmonic components, e.g. the second harmonic B_2. Hence if two transistors are matched for minimum second-harmonic content, they will produce minimum even-order harmonics (see Section 8).

The variation of the second harmonic with the out-of-balance ratio n was calculated using expression (4), and the results obtained are given in Fig. 1.

Since the paper is concerned only with distortion due to mis-matching in the transistors, the transformers are assumed to be
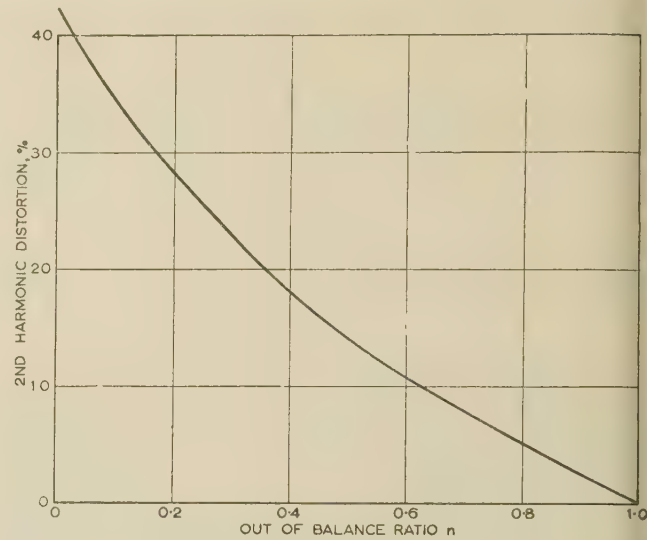
Fig. 1.—Variation of second-harmonic distortion with out-of-balance ratio n.

perfectly matched and the effects of transistor parameters are con-sidered separately. These parameters are

(a) Input resistance.
(b) Current gain factor and its variation with emitter current.

(2.1) Mismatch of Input Circuit Resistance

The voltage developed across the load R_L is given by

$$v_L = \frac{A_i R_L}{R_{in} + R_S} v_s$$
$$= G v_s \quad . \quad . \quad . \quad . \quad . \quad ($$

where A_i is the current gain of the amplifier (see Fig. 2).
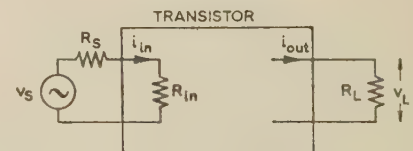
Fig. 2.—Nomenclature used in referring to each half of a push-pu amplifier.

Current gain: $Ai_1 = i_{out}/i_{in}$.
Voltage gain: $G_1 = v_L/v_s$.

In this Section it is considered that R_L, A_i, R_S and v_s are equa for the two halves of the amplifier, and distortion due to mis match of R_{in} only is considered.

(2.1.1) Common-Emitter Amplifier.

The input resistance of a transistor in the common-emitte connection at low frequencies is given[2] by

$$R_{in} \simeq r_{b0} + r_{b1} + r_e\left(\frac{1}{1 - \alpha}\right) \quad . \quad . \quad . \quad ($$

when r_c is large compared with r_e, r_b, r_{b0} and R_L.

The current in the input circuit is given by

$$i_s = \frac{v_s}{R_{in} + R_S}$$
$$= \frac{v_s}{R_S + r_{b0} + r_{b1} + r_e/(1 - \alpha)} \quad . \quad . \quad ($$

The emitter resistance r_e is given by $r_e \simeq \frac{kT}{I_e} \simeq \frac{25}{I_e}$ ohms, where I_e is in milliamperes. Also r_{b1} decreases in a similar manner with increasing emitter currents. In power amplifiers the emitter currents are sufficiently high to make r_e and r_{b1} negligible compared with r_{b0} and R_S.

Hence
$$i_s \simeq \frac{v_s}{R_S + r_{b0}} \quad \cdots \quad (8)$$

Therefore the principal cause of distortion in this case is due to r_{b0}, and from eqns. (5) and (8),
$$n = \frac{R_S + r_{b01}}{R_S + r_{b02}} \quad \cdots \quad (9)$$

Substituting for n in eqn. (4) the percentage qth-harmonic distortion is given by
$$\left(\frac{r_{b02} - r_{b01}}{2R_S + r_{b01} + r_{b02}}\right)B_q \times 100 \quad \cdots \quad (10)$$

If the base-resistance ratio r_{b01}/r_{b02} is given by m, the percentage qth-harmonic distortion is given by
$$\left[\frac{r_{b02}(1 - m)}{2R_S + r_{b02}(1 + m)}\right]B_q \times 100 \quad \cdots \quad (11)$$

The variation of distortion with m for particular values of R_S/r_{b01} is given in Fig. 3. Also shown in Fig. 3 are the distor-
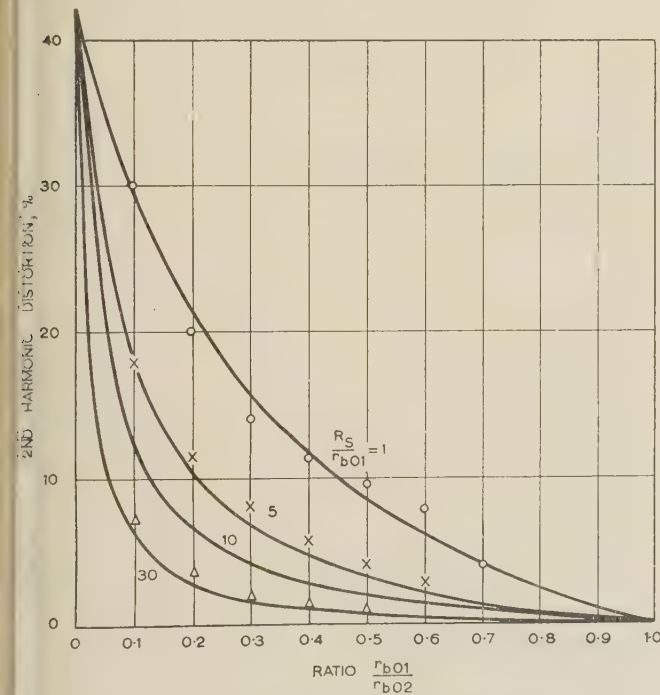
Fig. 3.—Variation of second-harmonic distortion with the ratio r_{b01}/r_{b02} for various values of R_S/r_{b01}.

tions that were obtained in a push-pull amplifier (common emitter). It can be seen that they are in very good agreement with the calculated values.

If a spread in r_{b0} of 45 to 90 ohms is assumed, the worst possible case of mismatch in r_{b0} is $0\cdot5$. It can be seen from Fig. 3 that, with this mismatch and a generator resistance of 1.5 kilohms (approximately 30 times greater than r_{b01}), the second-harmonic distortion is of the order of 1%, whereas if R_S is of the same order as r_{b01}, it is $8\cdot5\%$.

(2.1.2) Common-Base Amplifier.

The input resistance of a transistor in the common-base connection at low frequencies is approximately given by
$$R_{in} \simeq r_e + (r_{b0} + r_{b1})(1 - \alpha)$$

provided that the load resistance R_L is small compared with r_c. Hence, if the input signal v_s is fed from a generator resistance R_S, the input signal current is given by
$$i_s = \frac{v_s}{R_S + r_e + (r_{b0} + r_{b1})(1 - \alpha)}$$
$$\simeq \frac{v_s}{R_S + r_{b0}(1 - \alpha)} \quad \cdots \quad (12)$$

In this case, the effect of mismatch in the base resistance is reduced because of the multiplication factor $(1 - \alpha)$.

(2.1.3) Common-Collector Amplifier.

In the common-collector amplifier the input resistance is given by
$$R_{in} \simeq r_{b0} + r_{b1} + \frac{r_e + R_L}{1 - \alpha}$$

where R_L is small compared with $r_c(1 - \alpha)$. Thus
$$i_s \simeq \frac{v_s}{R_S + r_{b0} + r_{b1} + \dfrac{r_e + R_L}{1 - \alpha}} \quad \cdots \quad (13)$$

From this expression it is seen that the base resistance is swamped by the term $R_L/(1 - \alpha)$, and thus distortion due to mismatch of the base resistance is negligible.

(2.2) Mismatch of Current Gain Factor

Differences in the current gain factors of transistors can be classified in two ways. The mean gain for a given input signal voltage may not be the same, and, as seen from Fig. 4, the slope of the α/I_E curve may also differ.

(2.2.1) Mismatch of the Average Current Gain Factor.

Let the average current gain factors for a particular input signal be denoted by $\bar{\alpha}_1$ and $\bar{\alpha}_2$ for the two transistors. If the slopes of the α/I_e curves for the two transistors are such that α_1/α_2 is constant for all values of I_e, the distortion can be calculated using expression (4) as before.

Fig. 5 shows the second-harmonic distortion of a common-emitter amplifier as a function of the average current-gain ratio of the two transistors, $\bar{\alpha}_{1cb}/\bar{\alpha}_{cb2}$, calculated from expression (4). Fig. 5 also gives the measured values of second-harmonic distortion obtained with transistors having different values of $\bar{\alpha}_{cb}$. These results are in close agreement with the calculated values.

From Fig. 5 it can be seen that, in order to obtain less than $2\cdot5\%$ second-harmonic distortion, the gain factors $\bar{\alpha}_{cb}$ must be matched to within 10%.

(2.2.2) Mismatch of Slope β of the Current Gain Factor.

Typical variations of the current gain factors α and α_{cb} (common base and common emitter, respectively) with emitter current are shown in Figs. 4(a) and 4(b). It is seen from Fig. 4(a) that a close approximation to the variation of α is given[1] by
$$\alpha = \alpha_z - \beta I_e \quad \cdots \quad (14)$$

To a first approximation the mean value of α is given by
$$\bar{\alpha} = \frac{2\alpha_z - \beta I_{ep}}{2} \quad \cdots \quad (15)$$
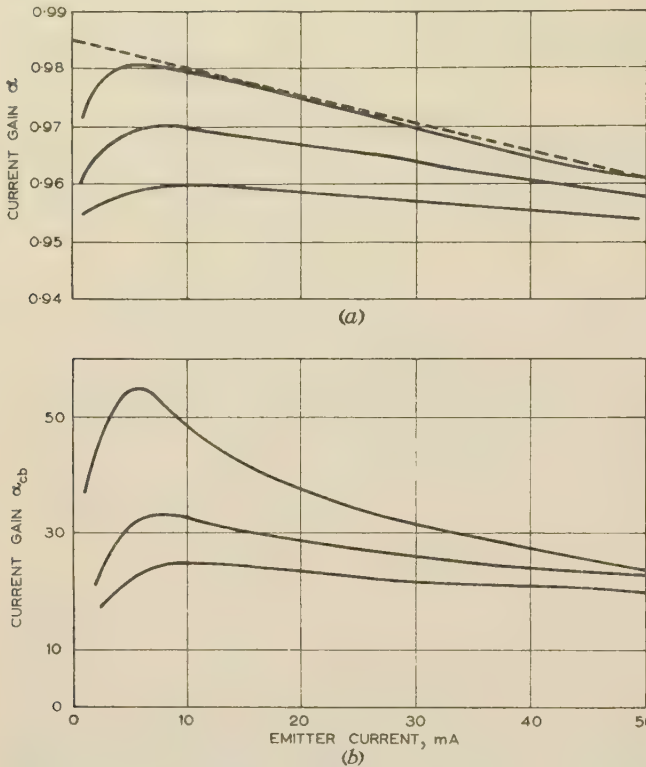
Fig. 4.—Variation of current gain factor with emitter current for three low-power transistors.

(a) α plotted against I_e.
···· $\alpha = \alpha_z \beta I_e$, where $\alpha_z = 0.985$ and $\beta = 0.0005$ per mA.
(b) α_{cb} plotted against I_e.

In the common-emitter configuration the mean current gain $\bar{\alpha}_{cb}$ is given by

$$\bar{\alpha}_{cb} = \frac{\bar{\alpha}}{1 - \bar{\alpha}}$$

$$= \frac{1}{2}\left(\frac{2\alpha_z - \beta I_{ep}}{1 - \alpha_z + \frac{\beta I_{ep}}{2}} \right) \quad . \quad . \quad . \quad (1)$$

If the two transistors have the same value of α_z but different slopes, β_1 and β_2, the ratio n may still be expressed approximately as the ratio of $\bar{\alpha}_{cb1}$ to $\bar{\alpha}_{cb2}$. Substituting for n in expression (4) the percentage second-harmonic distortion is given by

$$\frac{1 - \bar{\alpha}_{cb1}/\bar{\alpha}_{cb2}}{1 + \bar{\alpha}_{cb1}/\bar{\alpha}_{cb2}} B_2 \times 100 \quad . \quad . \quad . \quad (1$$

From expressions (16) and (17), the second-harmonic distortion

$$\frac{(2\alpha_z - \beta_2 I_{ep})\left(1 - \alpha_z + \frac{\beta_1 I_{ep}}{2}\right) - (2\alpha_z - \beta_1 I_{ep})\left(1 - \alpha_z + \frac{\beta_2 I_{ep}}{2}\right)}{(2\alpha_z - \beta_2 I_{ep})\left(1 - \alpha_z + \frac{\beta_1 I_{ep}}{2}\right) + (2\alpha_z - \beta_1 I_{ep})\left(1 - \alpha_z + \frac{\beta_2 I_{ep}}{2}\right)}$$

$$B_2 \times 100 \quad . \quad (1$$

From this expression it is found that the distortion due to different values of β is a function of peak emitter current, the current-gain slope β, α_z and the slope ratio β_1/β_2. Figs. 6, 7, and 8 show the percentage second-harmonic distortion as a function of β_1/β_2 for typical values of I_{ep}, β, and α_z as calculated from eqn. (18). For these calculations the slope of β_1 was less than that of β_2, and the resultant expression for the distortion has negative sign.

The sign obtained from the application of expression (4) important because some of the distortions do, in fact, cance Thus in a practical amplifier the two transistors are likely t have different values of α_z. If $\alpha_{z1} < \alpha_{z2}$ the sign of the di

Fig. 5.—Variation of second-harmonic distortion with the ratio of current gains $\alpha_{cb1}/\alpha_{cb2}$.
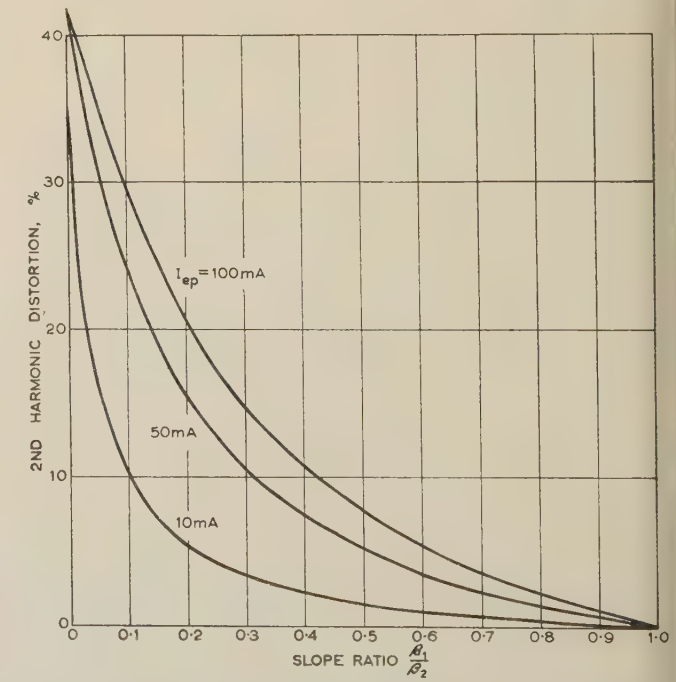
Fig. 6.—Variation of second-harmonic distortion with slope ratio for various values of I_{ep}.

$\alpha_z = 0.975$.
$\beta_1 = 0.00036$.
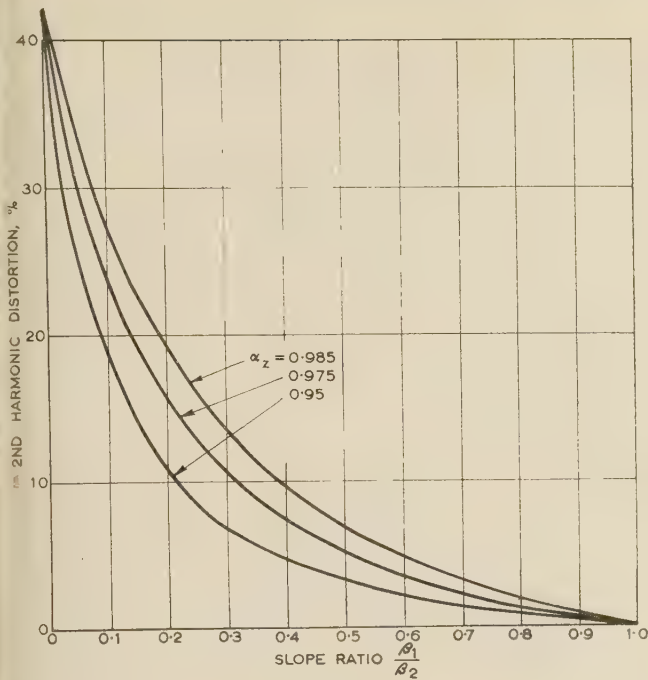
Fig. 7.—Variation of second-harmonic distortion with slope ratio for various values of α_z.
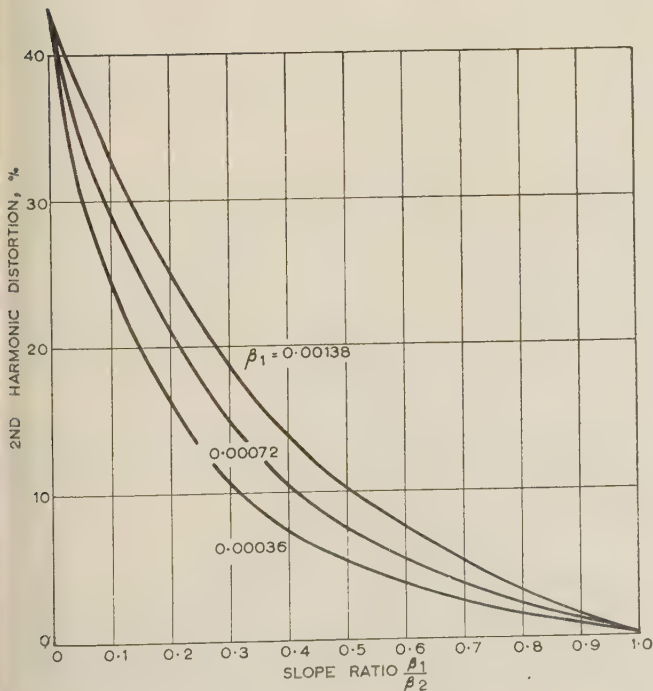
$$I_{ep} = 50\,\text{mA.}$$
$$\beta_1 = 0.000\,36.$$

tortion is positive, which means that the distortion due to differences in β_1 will be reduced if $\beta_1 < \beta_2$. This is an important consideration, since units having high values of α_z usually have high values of β; this may be seen from Figs. 4(a) and 4(b).

Owing to the number of variables involved, it is difficult to measure the distortion caused by differences in β. However, the effect is relatively small; a difference of 30% in the values of β would give less than 3% second-harmonic distortion.

(2.3) Application of the Results obtained in Sections 2.1 and 2.2

In this Section the effects of r_{b0}, $\bar{\alpha}$ and β are considered and their relative importance is assessed. It is shown that the normal production measurement of current gain factor at low emitter currents does not offer a satisfactory basis on which to match transistors for push-pull power amplifiers.

(2.3.1) Base Resistance.

For the particular type of transistor considered it was found that r_{b0} varied over the range 45–90 ohms. In common-emitter amplifiers the source resistance is, typically, as high as 1 000 ohms (in order to reduce cross-over distortion).[1] If two limit transistors are used (i.e. r_{b0} is 45 ohms for one and 90 ohms for the other) the second-harmonic distortion due to unequal values of r_{b0} is less than 2%.

(2.3.2) Current Gain Factor.

A spread of the order of 2 : 1 is obtained for $\bar{\alpha}_{cb}$. On its own this would give rise to a maximum second-harmonic distortion of about 15%. However, for two transistors with limit values of $\bar{\alpha}_{cb}$, the difference in β is such as to reduce the distortion by about 3% for a peak current swing of 30 mA; if the peak current is increased the distortion is further reduced, as may be seen from Fig. 11.

Fig. 8.—Variation of second-harmonic distortion with slope ratio for various values of β_1.
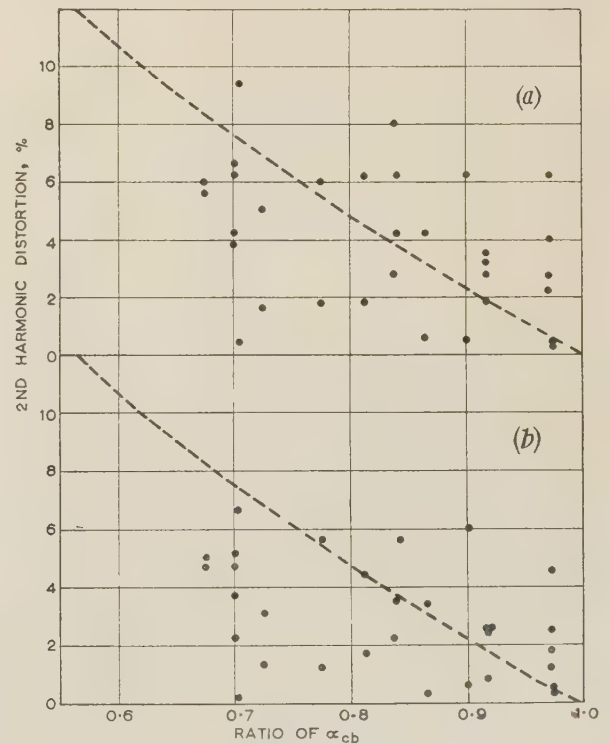
$$I_{ep} = 50\,\text{mA.}$$
$$\alpha_z = 0.975.$$

Fig. 9.—Variation of second-harmonic distortion with the ratio of α_{cb} (measured at $I_e = 1\,\text{mA}$).

1·5-kilohm source resistance.
- - - - Calculated value.
(a) $I_{ep} = 30\,\text{mA.}$
(b) $I_{ep} = 50\,\text{mA.}$

(2.3.3) Matching of Transistors for Push-Pull Operation.

To assess the combined effects of mismatch in $\bar{\alpha}$, β and r_{b0}, a number of a particular type of transistor were paired at random and operated in a common-emitter push-pull amplifier. The second-harmonic distortion was measured at peak emitter currents of 30 and 50 mA and plotted as a function of the ratio of various current-gain measurements. The results are shown in Figs. 9, 10 and 11. It is seen that, in all cases, the distortion is less than 10%.

Fig. 9 shows the results obtained using the current gain factors measured at emitter currents of 1 mA. It is seen that there is no correlation between these and the results theoretically predicted on the basis of Section 2.2. The reason can be seen from Fig. 4(b), in which it is shown that this current gain factor departs substantially from α_z and hence from $\bar{\alpha}_{cb}$, since α is greatly affected by the transistor surface conditions in this low current range.

The variation of the second-harmonic distortion with the α_{cb}
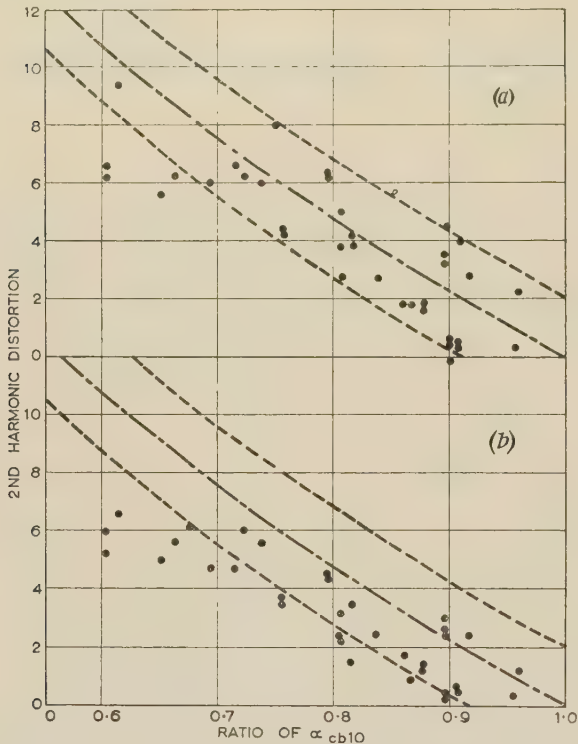
Fig. 10.—Variation of second-harmonic distortion with the ratio of α_{cb10} at $I_e = 10$ mA.

Source resistance = $1 \cdot 5$ kΩ.
= = = Indicates $\pm 2\%$ spread due to r_{b0}.
— — — Calculated value.
(a) $I_{ep} = 30$ mA.
(b) $I_{ep} = 50$ mA.

at $I_e = 10$ mA is shown in Fig. 10. It is seen that there is a considerable spread about the calculated value.

The range in values of r_{b0} for the transistors used (45–90 ohms), and the source resistance of 1 000 ohms, would result in a possible spread of $\pm 2\%$ in the second-harmonic distortion, which is confirmed by the experimental results. As mentioned earlier, transistors having a high value of α_{cb} also have a high value of β; this accounts for the distortion being lower at peak currents of 50 mA than at 30 mA, and lower than the calculated value for large degrees of mismatch.

One method of matching is to measure the gain of the amplifier under conditions which closely resemble the actual operating conditions, the only difference being that a d.c. input signal is
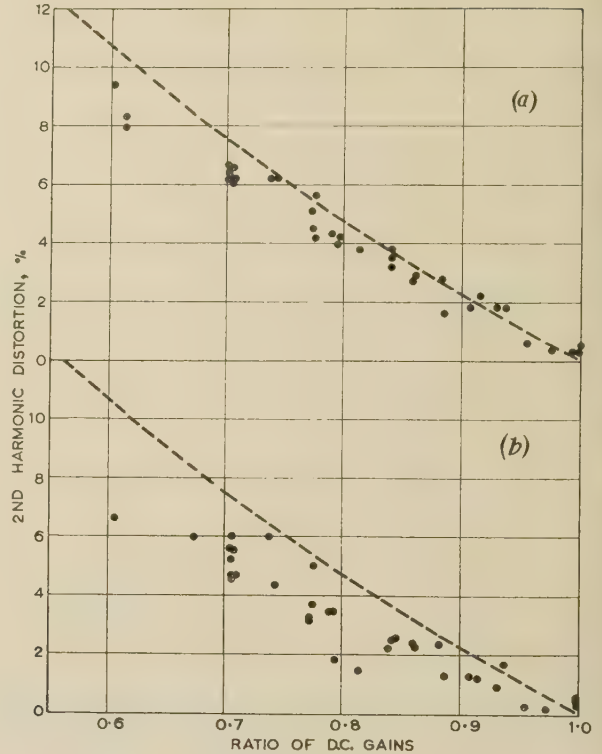
Fig. 11.—Variation of second-harmonic distortion with the ratio of gain (d.c. measurement).

Source resistance $1 \cdot 5$ kΩ.
— — — Calculated.
(a) $I_{ep} = 30$ mA.
(b) $I_{ep} = 50$ mA.

used instead of an a.c. one. A voltage is applied through a typical source resistance (e.g. 1 kilohm for a low-power transistor) to the base of the transistor, and the collector current is measured; the collector load is also chosen to represent a practical case. Fig. 11 shows the calculated and measured distortion, using the gains obtained by this experiment and applying the method discussed in Section 2.2.

It is seen that good agreement is obtained, and this fairly simple method of matching is therefore satisfactory. It is important to note that transistors matched for one set of operating conditions may not be matched for another; matching for all possible circuit conditions is cumbersome, since it involves independent matching of α_z, β and r_{b0}.

(3) MATCHING CONSIDERATIONS AT HIGH AUDIO FREQUENCIES

In this Section only the common-emitter amplifier is considered, since frequency effects are more pronounced in this arrangement than in the common-base or common-collector types. The main factors to be considered are:

(a) Collector capacitance.
(b) Variation of current gain factor with frequency.

(3.1) Collector Capacitance

The effective output capacitance C_{out} can be high in the common-emitter amplifier. If R_S is large compared with R_{in},

$$C_{out} \simeq \frac{\alpha}{1 - \alpha} C_c$$

where C_c is the capacitance of the collector-base p–n junction.

Since $C_c \propto V_c^{-\frac{1}{2}}$, the value of C_{out} can be high at the point corresponding to minimum collector voltage, i.e. high current. $1/(\omega C_{out})$ is comparable with R_L, the bandwidth of the amplifier will be limited by C_{out}. A difference in the values of C_{out} for the two transistors in a push-pull amplifier (owing to differences in α or C_c) can result in even-order harmonic distortion.

In most power-output stages R_L is very low compared with $1/(\omega C_{out})$ over the audio-frequency range, and the effects of C_{out} can therefore be ignored. However, when higher-power transistors having high collector capacitance are used for handling comparatively low output powers, C_{out} can have a significant effect.

(3.2) Variation of Current Gain with Frequency

The common-base current gain factor, α, varies with frequency according to the expression

$$\alpha = \frac{\alpha_0}{1 + jK} \quad . \quad . \quad . \quad . \quad (19)$$

where α_0 = Value of α at low frequencies.
$K = f/f_\alpha$, f being the operating frequency and f_d the frequency at which $|\alpha| = 0\cdot7\alpha_0$.

In the common-emitter amplifier the current gain factor α_{cb} is given by

$$\alpha_{cb} = \frac{\alpha}{1 - \alpha}$$

$$\alpha_{cb} = \frac{\dfrac{\alpha_0}{1 + jK}}{1 - \dfrac{\alpha_0}{1 + jK}}$$

hence

$$= \frac{\alpha_{cb0}}{1 + j\dfrac{K}{1 - \alpha_0}} \quad . \quad . \quad . \quad . \quad (20)$$

The cut-off frequency $f_{\alpha cb}$, i.e. the frequency at which $|\alpha_{cb}| = 0\cdot7\alpha_0/(1 - \alpha_0)$, is given by

$$f_{\alpha cb} = f_\alpha(1 - \alpha_0) \quad . \quad . \quad . \quad . \quad (21)$$

It has been verified experimentally that $f_{\alpha cb}$ varies with frequency according to eqn. (20) up to frequencies of at least $5f_{\alpha cb}$.

(3.2.1) Variation of Amplifier Gain with Frequency.

A common-emitter amplifier working from a source R_S into a load R_L is considered. The voltage gain G (as defined in Section 2) is given by

$$G = \frac{\alpha_{cb}R_L}{R_S + R_{in}} \quad . \quad . \quad . \quad . \quad (22)$$

From the expression given in Section 2, the input resistance R_{in} can be written as

$$R_{in} = r_{b0} + \frac{R_E}{1 - \alpha}$$

where R_E is the total resistance (i.e. r_e plus any added resistance) in the emitter circuit.

Therefore

$$G = \frac{\alpha_{cb}R_L}{R_S + r_{b0} + \dfrac{R_E}{1 - \alpha}} \quad . \quad . \quad . \quad (23)$$

The value of G will decrease with frequency owing to α_{cb} [see

eqn. (20)] and its modulus will decrease to $0\cdot7G_0$ (G_0 being its low-frequency value) at a frequency f_g, given by

$$f_g = f_\alpha\left(1 - \alpha_0\frac{R_S + r_{b0}}{R_S + r_{b0} + R_E}\right) \quad . \quad . \quad . \quad (24)$$

Since $(R_S + r_{b0})/(R_S + r_{b0} + R_E)$ is always less than unity, f_g is greater than $f_{\alpha cb}$ [which is equal to $(1 - \alpha_0)f_\alpha$].
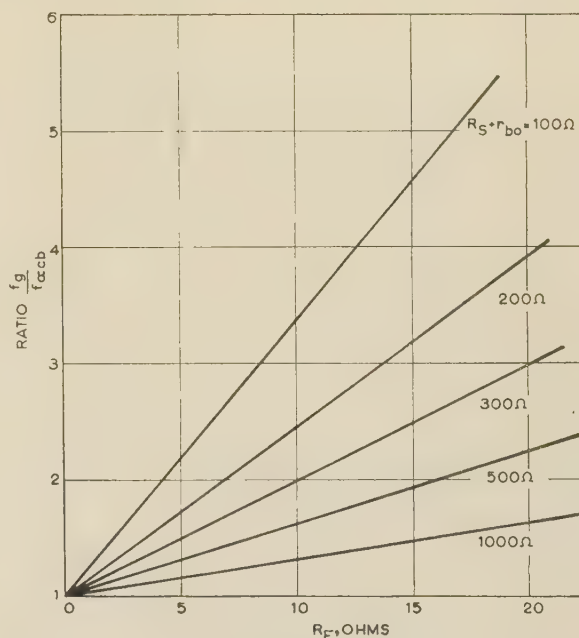
Fig. 12.—Variation of f_g with emitter resistance R_E for various source resistances.
$\alpha_0 = 0\cdot97$.

Fig. 12 shows the effect of the term $(R_S + r_{b0})/(R_S + r_{b0} + R_E)$; it compares f_g and $f_{\alpha cb}$ for various values of R_E and $(R_S + r_{b0})$. Since the difference between f_g and $f_{\alpha cb}$ for a given value of $(R_S + r_{b0})/(R_S + r_{b0} + R_E)$ increases as α_0 increases, the results in Fig. 12 have been taken for $\alpha_0 = 0\cdot97$.

Thus improvement can be obtained in the following two ways:

(a) Decrease $(R_S + r_{b0})$.
(b) Increase R_E.

It should be realized that this is merely a method of increasing the effective bandwidth of G at the expense of decreasing the gain at very low frequencies. Therefore the extent to which (a) and (b) may be applied depends on the decrease in gain that can be tolerated; a number of other limitations also arise.

The minimum value of R_S in most applications is limited by origin distortion considerations,[1] and therefore for low-power transistors R_S must, in general, be greater than 500 ohms; with no addition to R_E this value of R_S does not greatly increase f_g above $f_{\alpha cb}$.

R_E may be increased by adding external resistance in the emitter circuit. If $(R_S + r_{b0}) = 500$ ohms and $R_E = 10$ ohms, f_g is increased to approximately $1\cdot7f_{\alpha cb}$. However, the addition of external emitter resistance decreases the efficiency of the stage.

(3.3) Effects of Mismatch of Cut-Off Frequency

It may be assumed that the base-current input signals of a common-emitter amplifier are of a half sine-wave form (i.e. they consist of a fundamental and even-order harmonics) and the two halves are identical.

These are operated upon by G, as given by eqn. (22). The output voltage across the load R_L is thus given by

$$v_L = G_1 \left[A \frac{1}{\sqrt{(1 + H_{11}^2)}} \sin(\omega t - \phi_{11}) \right.$$

$$\cdot AB_2 \frac{1}{\sqrt{(1 + H_{12}^2)}} \cos(2\omega t - \phi_{12}) - \ldots$$

$$\left. - ABq \frac{1}{\sqrt{(1 + H_{1q}^2)}} \cos(q\omega t - \phi_{1q}) \right]$$

$$+ G_2 \left[A \frac{1}{\sqrt{(1 + H_{21}^2)}} \sin(\omega t - \phi_{21}) \right.$$

$$+ AB_2 \frac{1}{\sqrt{(1 + H_{22}^2)}} \cos(2\omega t - \phi_{22}) + \ldots$$

$$\left. + AB_q \frac{1}{\sqrt{(1 + H_{2q}^2)}} \cos(q\omega t - \phi_{2q}) \right] \quad . \quad . \quad . \quad (25)$$

where H is the ratio f_h/f_g and f_h is the frequency considered (i.e. the frequency of the fundamental or a harmonic). The suffixes denote the transistor and the harmonic, e.g. H_{24} is the ratio f_h/f_g in the second transistor for the fourth harmonic. ϕ is equal to arc tan H.

With the substitution $\sin \phi = H/\sqrt{(1 + H^2)}$ and $\cos \phi = 1/\sqrt{(1 + H^2)}$ the fundamental is given by

$$A \left[\left(\frac{G_1}{1 + H_{11}^2} + \frac{G_2}{1 + H_{21}^2} \right) \sin \omega t \right.$$

$$\left. - \left(\frac{G_1 H_{11}}{1 + H_{21}^2} + \frac{G_2 H_{21}}{1 + H_{21}^2} \right) \cos \omega t \right]$$

$$= A \sqrt{\left[\left(\frac{G_1}{1 + H_{11}^2} + \frac{G_2}{1 + H_{21}^2} \right)^2 \right.}$$

$$\left. + \left(\frac{G_1 H_{11}}{1 + H_{11}^2} + \frac{G_2 H_{21}}{1 + H_{21}^2} \right)^2 \right] \sin(\omega t - \theta_1) \quad . \quad (26)$$

where θ_1 is the phase shift and is given by

$$\theta_1 = \text{arc tan} \left(\frac{\dfrac{G_1 H_{11}}{1 + H_{11}^2} + \dfrac{G_2 H_{21}}{1 + H_{21}^2}}{\dfrac{G_1}{1 + H_{11}^2} + \dfrac{G_2}{1 + H_{21}^2}} \right) \quad . \quad . \quad (27)$$

The second harmonic is given by

$$AB_2 \left[\left(\frac{G_1}{1 + H_{12}^2} - \frac{G_2}{1 + H_{22}^2} \right) \cos 2\omega t \right.$$

$$\left. + \left(\frac{G_1 H_{12}}{1 + H_{12}^2} - \frac{G_2 H_{22}}{1 + H_{22}^2} \right) \sin 2\omega t \right]$$

$$= AB_2 \sqrt{\left[\left(\frac{G_1}{1 + H_{12}^2} + \frac{G_2}{1 + H_{22}^2} \right)^2 \right.}$$

$$\left. + \left(\frac{G_1 H_{12}}{1 + H_{12}^2} - \frac{G_2 H_{22}}{1 + H_{22}^2} \right)^2 \right] \cos(2\omega t - \theta_2) \quad . \quad (28)$$

where θ_2 is the phase shift and is given by

$$\theta_2 = \text{arc tan} \left(\frac{\dfrac{G_1 H_{12}}{1 + H_{12}^2} - \dfrac{G_2 H_{22}}{1 + H_{22}^2}}{\dfrac{G_1}{1 + H_{12}^2} - \dfrac{G_2}{1 + H_{22}^2}} \right) \quad . \quad . \quad (29)$$

Thus from expressions (26) and (28) the percentage second harmonic distortion is given by

$$100 B_2 \sqrt{\left[\frac{\left(\dfrac{G_1}{1 + H_{12}^2} - \dfrac{G_2}{1 + H_{22}^2} \right)^2 + \left(\dfrac{G_1 H_{12}}{1 + H_{12}^2} - \dfrac{G_2 H_{22}}{1 + H_{22}^2} \right)^2}{\left(\dfrac{G_1}{1 + H_{11}^2} + \dfrac{G_2}{1 + H_{21}^2} \right)^2 + \left(\dfrac{G_1 H_{11}}{1 + H_{11}^2} + \dfrac{G_2 G_{21}}{1 + H_{21}^2} \right)^2} \right]}$$

$$. \quad . \quad . \quad (30)$$

Assuming that matching at low frequency has been carried ou $G_1 \simeq G_2$. Hence the percentage second-harmonic distortion is

$$100 B_2 \sqrt{\left[\frac{\left(\dfrac{1}{1 + H_{21}^2} - \dfrac{1}{1 + H_{22}^2} \right)^2 + \left(\dfrac{H_{12}}{1 + H_{12}^2} - \dfrac{H_{22}}{1 + H_{22}^2} \right)^2}{\left(\dfrac{1}{1 + H_{11}^2} + \dfrac{1}{1 + H_{21}^2} \right)^2 + \left(\dfrac{H_{11}}{1 + H_{11}^2} + \dfrac{H_{21}}{1 + H_{21}^2} \right)^2} \right]}$$

$$. \quad . \quad . \quad (31)$$

From eqn. (25) the phase shift and amplitude of any harmoni may be calculated by substituting the appropriate values of A and H for any given mismatch of f_g.

The variation of second-harmonic distortion with frequenc for various ratios of cut-off frequency were calculated usin eqn. (31) and are shown in Fig. 13. It can be seen that, with
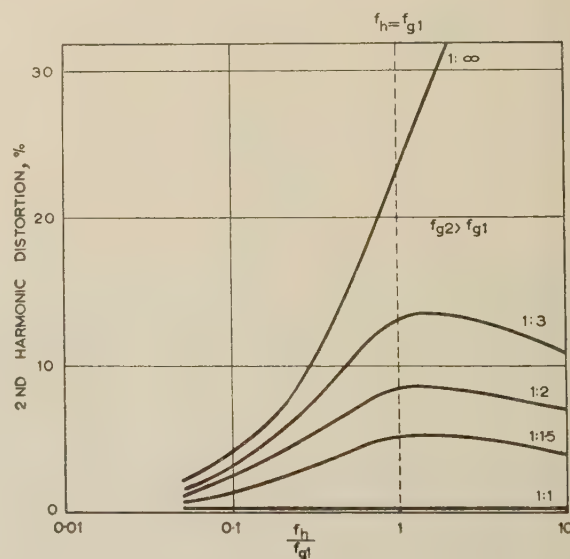
Fig. 13.—Variation of second-harmonic distortion with frequency fo various values of mismatch in f_g.

large mismatch of f_g, 4% distortion is obtained at one-tenth o the lower cut-off frequency. These curves can be used to obtain the higher-order harmonic distortion, since, for a give value of f_h/f_g, the percentage distortion is proportional to th values of B given in Section 2.

(3.4) Effects of the Spread in Transistor Characteristics

Most transistor characteristics show variation between units For example, at $V_c = -6$ volts and $I_e = 1$ mA, the following spread in characteristics was obtained with the type of transisto considered:

(i) f_α ... $0 \cdot 8$–$2 \cdot 8$ Mc/s
(ii) C_c ... 55–85 pF
(iii) α_{cb} ... 40–80
(iv) $f_{\alpha cb}$ ... 15–48 kc/s

The spread of $f_{\alpha cb}$ is less than would be expected from taking the limit cases of (i) and (ii). This is because transistors with a high current gain tend to have a high cut-off frequency.

C_c is sufficiently small in these transistors for its effect on power-amplifier performance to be ignored.

It is found that $f_{\alpha cb}$ varies with emitter-current bias in an inverse manner to α_{cb}; this is to be expected since f_α is substantially independent of emitter current. Therefore, the minimum value of $f_{\alpha cb}$ occurs at the maximum value of α_{cb}; in the transistor considered, this corresponds to an emitter current of about 10 mA.

Fifty pairs of transistors were matched for G to within $\pm 5\%$ at low frequency by the method described in Section 2. $f_{\alpha cb}$ was measured ($V_c = -6$ volts and $I_e = 10$ mA) for each transistor, and the results obtained for the individual pairs are plotted in Fig. 14. All pairs lie within a $f_{\alpha cb1}/f_{\alpha cb2}$ range of less
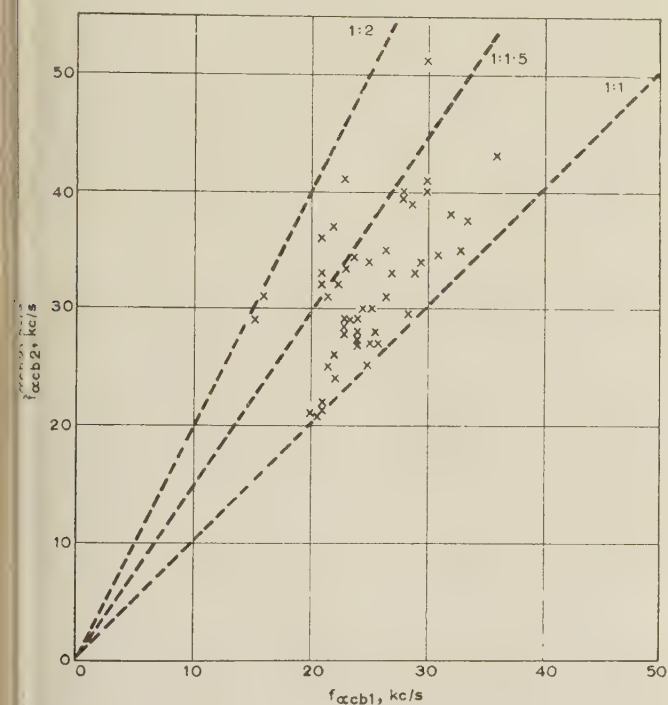
Fig. 14.—Mismatch spread of $f_{\alpha cb1}/f_{\alpha cb2}$ for 50 matched pairs of the type of transistor considered.

than 2 : 1; 42 pairs are within 1·5 : 1. Applying the results of Section 3.3, it would be expected that the distortion at 10 kc/s (in a class-B push-pull power amplifier) would be less than 5% for 43 pairs and less than 8% for all pairs. When the distortion was measured experimentally the results shown in Fig. 15 were obtained; in the amplifier the peak collector current was approximately 30 mA. It is seen that 44 pairs have distortions less than 3·5%, and the remainder are less than 5%.

The lower distortions obtained in the practical case can be attributed to the fact that the relevant cut-off frequency should be the mean value over the emitter-current range used rather than the small-signal value at $I_e = 10$ mA used in the calculation.

(3.5) Reduction of Distortion at High Frequency by means of Matching

A second-harmonic distortion of less than 5% at 10 kc/s is sufficiently low for most applications. In this case, the results described in Section 3.4 indicate that matching of low-frequency gain is adequate for the type of transistor considered.
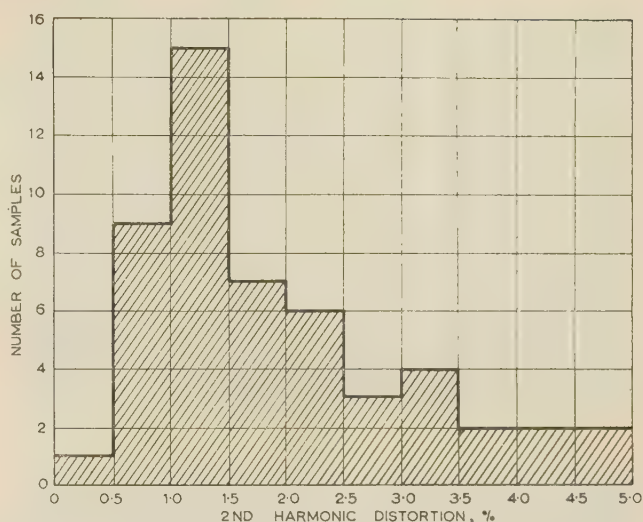
Fig. 15.—Histogram of second-harmonic distortion measured at 10 kc/s with 50 matched pairs of the type of transistor considered.

However, if lower distortion is required from amplifiers using these transistors, or if other types of transistor having lower values of $f_{\alpha cb}$ are used, some further matching is necessary. A modification of the low-frequency matching method to higher frequency (i.e. by using an a.f. input signal) is not suitable for high-frequency matching, since it would be an amplitude matching method and would not take phase differences into account. The accuracy of such a method is increased if the measurement is made at a frequency greater than the cut-off frequency. However, this necessitates a knowledge of the range of cut-off frequencies of the device. A better method is to match by means of the cut-off-frequency measurement, since this gives both the amplitude and phase differences.

A measurement of f_g may be made by using the matching arrangement suggested in Section 2, but with the base supplied from a variable-frequency source instead of with direct current. A measurement of mean collector current is made at a suitable low frequency, and the frequency is then increased until the output current has decreased by 3 dB.

(4) CARRIER-STORAGE EFFECTS

When the frequency of operation of a class-B amplifier is increased the output waveform can become very distorted because of carrier-storage effects. These occur even with a pair of transistors which are perfectly matched in all their parameters. An example of this distortion is shown in Fig. 16(a); this is an output waveform obtained at a frequency of 10 kc/s. It is seen that the distortion takes the form of a high-frequency 'ringing' superimposed upon the fundamental waveform. This is more clearly seen in Fig. 16(b), in which the fundamental waveform has been removed. The ringing occurs at the change-over from one half of the amplifier to the other.

The cause can be seen by examining the waveforms of one half of the amplifier. Fig. 17(a) shows the waveform obtained at the collector load of the amplifier, whereas Fig. 17(b) shows the waveform obtained by passing a half sine-wave through an RC network of the same frequency characteristics as the transistor. It is seen that a much sharper current cut-off is obtained from the amplifier than from the network. The reason can be seen in Fig. 17(c), which shows the base input current waveform. Thus, instead of the expected half sine-wave there is a reverse current pulse when the base is driven positively. This reverse pulse is

(a)

(b)
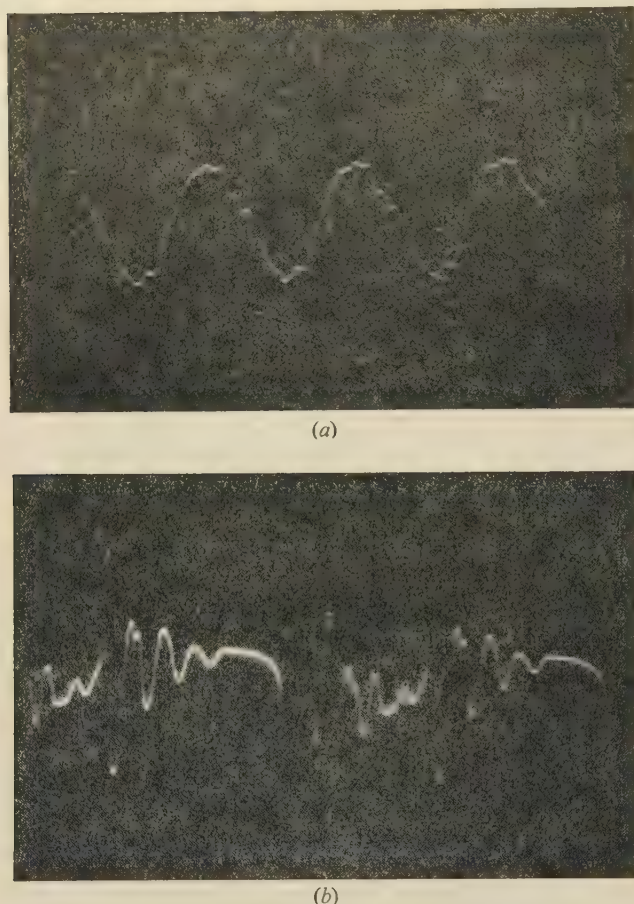
Fig. 16.—Waveforms of a class-B push-pull amplifier at 10 kc/s.

(a) Output waveform.
(b) Ringing of the waveform in (a) with the fundamental removed.

(a)

(b)

(c)

Fig. 17.—Waveforms of one half of the amplifier.

(a) Collector-current waveform.
(b) Effect of an RC network on a half sine wave.
(c) Base input-current waveform.

due to the minority carriers (holes in a p–n–p transistor) which are in the base region and which must be removed before the reverse input impedance becomes high.

If two waveforms of the form given in Fig. 17(b) are added together in the output circuit of a push-pull amplifier, a sine-wave output signal is obtained. However, adding together collector waveforms, as shown in Fig. 17(a), gives an output which consists of a sine wave with pulses superimposed on it at the cross-over points; the input waveform is similarly distorted. These pulses give rise to ringing in the input and output transformers.

The ringing frequency is determined by peaks in the transformer response curve. These occur at high frequencies owing to resonance between the leakage inductance and the self-capacitance of the transformers.[3] The Q-factor of the resonant circuit depends on the magnitude of the winding resistance and the transistor impedance relative to the leakage inductance. The damped oscillations given in Fig. 16(b) correspond to the input transformer having a Q-factor of 2 to 3. This can be reduced by increasing the resistance of the secondary winding.

In general, the ringing frequencies are well above the a.f. range, e.g. 50–60 kc/s in Fig. 16(b); however, they may give rise to unwanted signals which are fed back to other amplifier circuits in the equipment, e.g. the i.f. amplifier of a broadcast receiver.

The carrier-storage effect in junction transistors is dependent upon the thickness of the base region. Since f_α is also dependent upon the base thickness, the carrier storage is decreased as f_α is increased. The waveforms shown in Fig. 17(c) were obtained with a transistor having $f_\alpha = 0.8$ Mc/s, whereas Fig. 18(a) shows

the waveforms obtained using a similar transistor having $f_\alpha = 1.8$ Mc/s; the latter has a considerably lower carrier storage effect.

Fig. 18(b) shows the waveforms obtained with a higher frequency transistor. Little carrier-storage is exhibited. In general, medium-power transistors have a lower value of f_α, and as expected, they exhibit a larger carrier-storage effect.

The ringing effect can be eliminated by reducing the reverse current pulse. This can be achieved by connecting a low-hole storage diode in series with the base of each transistor. In the case of a low-power transistor, a point-contact diode can be used since these have a lower carrier storage than junction transistors. The diode is connected in series with the base, so that the current through the diode is considerably less than the emitter current, and this results in a small carrier-storage effect. The waveforms shown in Fig. 18(c) were obtained with such an arrangement [cf. Fig. 17(c)]. With higher-power transistors, a junction diode would have to be used because of the high base current; junction diodes exhibiting a low carrier storage are not readily available.

In addition, for thermal stability the d.c. gain of the amplifier should be as low as possible; this requires a low-d.c.-resistance base circuit. Thus the inclusion of a diode in series with the base decreases the thermal stability and increases the possibility of runaway.

An alternative method, which decreases the ringing effect, is

device, and a collector load resistance is included such that the maximum collector current is equal to the rated peak value. The input voltage step is such that the resultant collector current lies between 40 and 90% of the rated peak value.

A suitable measuring circuit for the type of transistor considered is shown in Fig. 19(a). Using this circuit, transistors can be matched on the basis of the measured collector current.
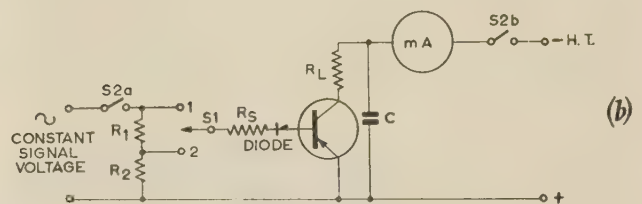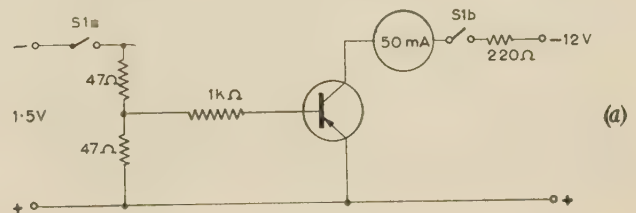
Fig. 19.—Transistor matching circuits.

    (a) L.F. matching.
        S1: 2-pole bias switch (normally open).
    (b) H.F. matching.
        S1: Bias switch (normally in position 1).
        S2: 2-pole bias switch (normally open).
        R_1/R_2 : 3/7.

To obtain the cut-off frequency f_g, required for high-frequency matching, the low-frequency circuit can be adapted as shown in Fig. 19(b). The base is now supplied with an alternating voltage, whose value is constant with frequency. A diode is inserted in series with the base to overcome carrier-storage effects and the meter is decoupled to the a.c. signal.

(6) CONCLUSIONS

Even-harmonic distortion can occur in a class-B push-pull amplifier owing to the mismatch of several of the transistor parameters.

At low frequencies, three parameters are important. Of these, the most important is the current gain factor. However, to obtain a low level of even-order harmonic distortion, all three parameters must be matched. This is difficult as a production technique; instead, a simple test circuit which simulates a typical amplifier arrangement can be used. This is found to give sufficiently well matched pairs of transistors.

At higher frequencies considerable harmonic distortion can arise owing to the mismatch of the common-emitter cut-off frequencies. This is very important if the operating frequency is close to the cut-off frequency of either of the transistors. Thus, with power transistors having a low-frequency response, some frequency matching is essential. For this reason it is highly desirable that the power-handling range of transistors having good high-frequency characteristics should be extended as far as possible by the application of efficient cooling systems.

The cross-over transients obtained at higher audio frequencies are caused by carrier-storage effects and cannot be overcome by matching. These transients do not seriously affect a.f. amplifiers but may be important in higher-frequency applications. Their

Fig. 18.—Waveforms of one half of the amplifier.

(a) A transistor with high cut-off operating at 10 kc/s.
(b) High-frequency transistor operating at 10 kc/s.
(c) The transistor used in Fig. 17(c) with a diode in the base circuit operating at ::c/s.
      (i) Base input-current waveform.
      (ii) Collector-current waveform.

use of bifilar secondary windings. This ensures a close uppling between the two windings and hence a low leakage ductance.[4]

(5) MATCHING

From Section 2 it is seen that a d.c. gain measurement is required for low-frequency matching, and a test circuit simulating typical amplifier arrangement is suggested. A direct voltage applied to the base via a typical source resistance. The tor supply voltage is adjusted to the rated value for the

effects can be reduced by circuit techniques, including the use of transformers having bifilar windings.

The transistor considered throughout the paper is the type GET4. The diodes used for Fig. 18(c) are type GEX54.

(7) REFERENCES

(1) HILBOURNE, R. A., and JONES, D. D.: 'Transistor Power Amplifiers', *Proceedings I.E.E.*, Paper No. 1861 R, April, 1955 (**102 B**, p. 763).

(2) SHEA, R. F. (Ed.): 'Principles of Transistor Circuits' (Wiley, 1953).

(3) TERMAN, F. E.: 'Radio Engineers Handbook' (McGraw-Hill, 1950).

(4) 'Transistors I' (RCA Laboratories, 1956).

(8) APPENDIX

(8.1) Second Harmonic Power Relative to the Total Harmonic Power

A sine wave of peak amplitude A is considered of which alternate half-waves are amplified by the factors G_1 and G_2, respectively ($G_1 > G_2$). The total power in the output load resistance, R_L, is given by

$$P_{tot} = \left\{ \frac{1}{2} \left[\frac{\sqrt{(2)}AG_1}{2} \right]^2 + \frac{1}{2} \left[\frac{\sqrt{(2)}AG_2}{2} \right]^2 \right\} \frac{1}{R_L}$$

$$= (G_1^2 + G_2^2) \frac{A^2}{4R_L}$$

The output waveform can be divided into its frequency components. Thus

$$P_{tot} = P_{dc} + P_1 + P_2 + P_4, \text{ etc.}$$
$$= P_{dc} + P_1 + P_h$$

For one half of the output wave (peak amplitude GA),

$$V = \frac{GA}{\pi} + \frac{GA}{2} \sin \omega t - \frac{GA}{2}(B_2 \cos 2\omega t + B_4 \cos 4\omega t, \text{ etc.})$$

Thus, adding the two halves of the output waveform which are of different amplitudes,

$$V_{dc} = (G_1 - G_2)\frac{A}{\pi}$$

Therefore
$$P_{dc} = (G_1 - G_2)^2 \frac{A^2}{\pi^2} \frac{1}{R_L}$$

$$V_1 = (G_1 + G_2)\frac{A}{2} \sin \omega t$$

Therefore
$$P_1 = (G_1 + G_2)^2 \frac{A^2}{8} \frac{1}{R_L}$$

$$V_2 = (G_1 - G_2)\frac{B_2 A}{2} \cos \omega t$$

Therefore
$$P_2 = (G_1 - G_2)^2 \frac{B_2^2 A^2}{8} \frac{1}{R_L}$$

$$V_h = (G_1 - G_2)\frac{A}{2}(B_2 \cos 2\omega t + B_4 \cos 4\omega t + \text{etc.})$$

Therefore
$$P_h = (G_1 - G_2)^2 \frac{A^2}{8} W \frac{1}{R_L}$$

where W is constant.

Now
$$P_h = P_{tot} - P_{dc} - P_1$$

Substitution gives
$$W = \frac{\pi^2 - 8}{\pi^2}$$

Therefore
$$\frac{P_2}{P_h} = \frac{B_2^2}{\frac{\pi^2 - 8}{\pi^2}}$$

But
$$B_2 = \frac{4}{3\pi}$$

Hence
$$\frac{P_2}{P_h} = \frac{16}{9(\pi^2 - 8)} \simeq \frac{8}{9}$$

Thus approximately 89% of the total harmonic power is in the second harmonic.

AN IMPROVED DESIGN PROCEDURE FOR THE MULTI-SECTION GENERALIZED MICROWAVE FILTER

By R. LEVY, M.A.

SUMMARY

A highly accurate expression for the Q-factor of a direct-coupled filter is derived which is analogous to that for an associated quarter-wave-coupled filter. Existing theories usually consider filters of either the maximally-flat or the Chebȳshev type, and it is shown directly, by transmission-line methods, that these are special cases of a generalized symmetrical filter. Design formulae are given for the bandwidth, pass-band tolerance, and the attenuation in the rejection band both for quarter-wave-coupled and direct-coupled generalized filters.

LIST OF SYMBOLS

$\left.\begin{array}{l} A, B, C, D \\ l, m, n \end{array}\right\}$ = Transfer matrix parameters.

Y = Admittance.

Z = Impedance.

$\theta \ (= 2\pi l/\lambda_g)$ = Electrical length of a cavity.

ϕ = Electrical length of a quarter-wave coupling.

R = Complex voltage insertion loss ratio.

$\left.\begin{array}{l} L \text{ (power ratio)} \\ \mathscr{L} \text{ (dB)} \end{array}\right\}$ = Power insertion loss.

ψ = Insertion phase shift.

jb = Normalized susceptance.

f = Frequency.

f_0 = Resonant, or mid-band, frequency.

$\Delta f = |f - f_0|$.

λ_g = Guide wavelength.

λ = Free-space wavelength.

Q = Loaded Q-factor.

$Q_{L(r)}$ = Loaded Q-factor for the rth cavity (in a maximally flat filter).

$Q_{L(1)}$ = Loaded Q-factor of each cavity (in a Chebȳshev filter).

Q_T = Overall loaded Q-factor.

Q_x, Q_y and Q_z = Loaded Q-factors for the respective cavities.

$x, y, z = 2Q_x\Delta f/f_0$, etc.

n = Number of cavities.

$U_{n(x)}$ = Rationalized Chebȳshev polynomial of the second kind of degree n.

$k = Q_x/Q_y$ (3 or 4 cavities).
$\ \ = Q_y/Q_z$ (5 or 6 cavities).

$l = Q_x/Q_z$ (5 or 6 cavities).

$X = 2l^2 + 2l - k$.

$Y = 2l - 2k + 1$.

$X' = kl^2 + 2l^2 + 2l - k$.

$Y' = l^2 + 2lk + 2l - k^2 - 2k + 1$.

y_1, z_1 = Values of y and z, respectively, corresponding to 3 dB points.

$y_T = 2Q_T\dfrac{\Delta f}{f_0}$ (3 or 4 cavities).

$z_T = 2Q_T\dfrac{\Delta f}{f_0}$ (5 or 6 cavities).

$\left.\begin{array}{l} L_T \text{ (power ratio)} \\ \mathscr{L}_T \text{ (dB)} \end{array}\right\}$ = Pass-band tolerance.

$\left.\begin{array}{l} L_R \text{ (power ratio)} \\ \mathscr{L}_R \text{ (dB)} \end{array}\right\}$ = Attenuation in the rejection band.

Q_{DC} = Q-factor of the direct-coupled filter.

$Q_{\lambda_g/4}$ = Q-factor of the associated quarter-wave coupled filter.

$$[x] = \begin{bmatrix} -x & j(1-x) \\ j(1+x) & -x \end{bmatrix}$$

(1) INTRODUCTION

A general synthesis procedure for the design of narrow-band direct-coupled filters has been developed by Hessel *et al.*[1] and by Riblet,[2] who have shown an approximate equivalence between quarter-wave and direct-coupled filters. The approximation used, applicable only to narrow-band filters, results in rather complicated expressions. However, a more exact treatment is possible which gives comparatively simple and more accurate design information, and provides a clearer physical insight into the relationship between the two types of filters.

(2) PRELIMINARY THEORY

The box in Fig. 1 represents any linear, passive network, and its external behaviour can be represented by the two linear relations

$$V_1 = AV_2 - B'I_2 \qquad . \quad . \quad . \quad . \quad (1)$$

$$I_1 = C'V_2 - DI_2 \qquad . \quad . \quad . \quad . \quad (2)$$
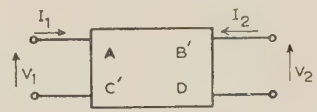
Fig. 1.—Representation of a network.

A and D are dimensionless, while B' and C' have the dimensions of impedance and admittance, respectively. Only three of the parameters are independent since the application of the reciprocity theorem gives

$$AD - B'C' = 1 \qquad . \quad . \quad . \quad . \quad (3)$$

The linear relations (1) and (2) are written in matrix form as follows:

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B' \\ C' & D \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix} \qquad . \quad . \quad . \quad (4)$$

where $\begin{bmatrix} A & B' \\ C' & D \end{bmatrix}$

is referred to as the *transfer matrix*.

The networks considered in the paper are composed of basic four-terminal networks as follows:

Shunt admittance Y . . . . $\begin{bmatrix} 1 & 0 \\ Y & 1 \end{bmatrix}$

Series impedance Z . . . . $\begin{bmatrix} 1 & Z \\ 0 & 1 \end{bmatrix}$

Line of electrical length θ of characteristic admittance unity . . . . $\begin{bmatrix} \cos \theta & j \sin \theta \\ j \sin \theta & \cos \theta \end{bmatrix}$
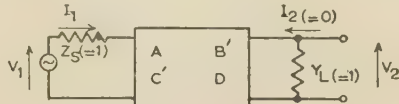
Fig. 2.—Derivation of insertion loss function.

Waveguide filters, etc., consist of combinations of these basic networks in cascade, and the transfer matrix of the complete network is obtained by multiplication of the component matrices in the appropriate order.

In Fig. 2 a generator is represented by a voltage source V_1 in series with a source impedance $Z_s = 1$, and the load is an admittance $Y_L = 1$, i.e. the generator and load are normally matched. They are separated by a 4-terminal network whose insertion loss and insertion phase shift are required as a function of the transfer parameters. The transfer matrix for the cascade connection of source, network and load is

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A & B' \\ C' & D \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

i.e. $\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} (A + D) + (B' + C') & B' + D \\ C' + D & D \end{bmatrix} \begin{bmatrix} V_2 \\ 0 \end{bmatrix}$

Therefore $\qquad \dfrac{V_1}{V_2} = (A + D) + (B' + C')$.

If R is the ratio of the voltage across the load when connected directly to the generator to that when connected through the network, it is seen that

$$R = \tfrac{1}{2}(A + D) + \tfrac{1}{2}(B' + C')$$

R gives the effect of insertion of the network in both amplitude and phase. Only lossless networks are considered in the paper, so that B' and C' may be replaced by the quantities jB and jC, where B and C are real and have the dimensions of resistance and conductance, respectively. The expression for R becomes

$$R = \tfrac{1}{2}(A + D) + \tfrac{1}{2}j(B + C) \quad . \quad . \quad . \quad . \quad (5)$$

The power insertion loss is

$$L = |R|^2 = 1 + \tfrac{1}{4}(A - D)^2 + \tfrac{1}{4}(B - C)^2 \quad . \quad (6)$$

and the insertion phase shift is

$$\psi = \arg (R) = \arctan \frac{B + C}{A + D} \quad . \quad . \quad . \quad (7)$$

In the particular case of symmetrical networks, i.e. those whose input and output terminals cannot be distinguished by external measurements, the parameters necessary to specify the four-terminal network are reduced to two by the additional relation

$$A = D \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and eqn. (6) can be simplified to give

$$L = 1 + \tfrac{1}{4}(B - C)^2 \quad . \quad . \quad . \quad . \quad (9)$$

(3) QUARTER-WAVE COUPLING SECTIONS

A microwave filter cavity consists of a section of waveguide terminated in obstacles of negligible thickness, which may be represented by pure susceptances. In order to obtain a filter with a 'flat-top' characteristic giving high rejection outside the pass band it is necessary to couple two or more cavities in cascade. In the case of quarter-wave-coupled filters the coupling sections consist of short lengths of waveguide. It is required to find the exact length of the coupling section between any two cavities, which may have different loaded Q-factors, irrespective of the shape of the overall filter characteristic, whether of maximally flat, Chebyshev, or of generalized type (Section 6). This length is one which gives perfect transmission between the cavities at the mid-band frequency.
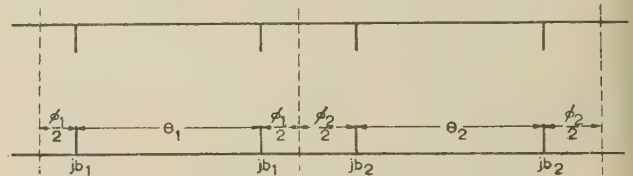
Fig. 3.—Section of a quarter-wave-coupled filter.

Fig. 3 shows part of a quarter-wave-coupled filter, consisting of two unequal cavities, specified by the subscripts 1 and 2. All susceptances considered in the paper are normalized with respect to the waveguide characteristic admittance. Thus the first cavity consists of two normalized susceptances jb_1 separated by an electrical distance θ_1. The transfer matrix for this cavity is

$$\begin{bmatrix} 1 & 0 \\ jb_1 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta_1 & j \sin \theta_1 \\ j \sin \theta_1 & \cos \theta_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jb_1 & 1 \end{bmatrix} = \begin{bmatrix} l_1 & jm_1 \\ jn_1 & l_1 \end{bmatrix} \quad (10)$$

where $\qquad l_1 = \cos \theta_1 - b_1 \sin \theta_1$
$$m_1 = \sin \theta_1$$
$$n_1 = 2b_1 \cos \theta_1 + (1 - b_1^2) \sin \theta_1.$$

The insertion loss for this single cavity is given by eqn. (6),

i.e. $\qquad L = 1 + \tfrac{1}{4}(m_1 - n_1)^2$
$$= 1 + b_1^2 (\cos \theta_1 - \tfrac{1}{2}b_1 \sin \theta_1)^2 \quad . \quad . \quad (11)$$

Hence a single cavity behaves like a filter with mid-band frequency given by the equation $m_1 = n_1$,

i.e. $\qquad \tan \theta_1 = \dfrac{2}{b_1} \quad . \quad . \quad . \quad . \quad . \quad (12)$

The transfer matrix for this cavity when enclosed by lines of length $\phi_1/2$ at either end is

$$\begin{bmatrix} \cos \dfrac{\phi_1}{2} & j \sin \dfrac{\phi_1}{2} \\ j \sin \dfrac{\phi_1}{2} & \cos \dfrac{\phi_1}{2} \end{bmatrix} \begin{bmatrix} l_1 & jm_1 \\ jn_1 & l_1 \end{bmatrix}$$

$$\begin{bmatrix} \cos \dfrac{\phi_1}{2} & j \sin \dfrac{\phi_1}{2} \\ j \sin \dfrac{\phi_1}{2} & \cos \dfrac{\phi_1}{2} \end{bmatrix} = \begin{bmatrix} l_1' & jm_1' \\ jn_1' & l_1' \end{bmatrix} \quad (13)$$

where
$$l_1' = l_1 \cos \theta_1 - \tfrac{1}{2}(m_1 + n_1) \sin \phi_1$$
$$m_1' = l_1 \sin \phi_1 + \tfrac{1}{2}(m_1 + n_1) \cos \phi_1 + \tfrac{1}{2}(m_1 - n_1)$$
$$n_1' = l_1 \sin \phi_1 + \tfrac{1}{2}(m_1 + n_1) \cos \phi_1 - \tfrac{1}{2}(m_1 - n_1)$$

ence the overall transfer matrix for two filter sections in cascade, as shown in Fig. 3, is

$$\begin{bmatrix} l_1' & jm_1' \\ jn_1' & l_1' \end{bmatrix}\begin{bmatrix} l_2' & jm_2' \\ jn_2' & l_2' \end{bmatrix} = \begin{bmatrix} l_1'l_2' - m_1'n_2' & j(l_1'm_2' + m_1'l_2') \\ j(n_1'l_2' + l_1'n_2') & l_1'l_2' - n_1'm_2' \end{bmatrix} \quad (14)$$

and the insertion loss for the two sections [eqn. (6)] is

$$L = 1 + \tfrac{1}{4}[m_1'n_2' - m_2'n_1']^2 + \tfrac{1}{4}[l_1'(m_2' - n_2') + l_2'(m_1' - n_1')]^2 \quad (15)$$

If the mid-band frequencies of the two cavities are made to coincide [eqn. (12)],

$$\tan\theta_1 = \frac{2}{b_1}, \tan\theta_2 = \frac{2}{b_2} \quad . \quad . \quad . \quad . \quad (16)$$

this will be the mid-band frequency of the cascade combination, with perfect transmission at this frequency, since eqn. (16) gives $m_1' = n_1'$, $m_2' = n_2'$, which makes L equal to unity in eqn. (15). The solutions of equations such as eqn. (16) are chosen so that θ lies in the range $\pi/2 < \theta < 3\pi/2$. Since the susceptances used in microwave filters are usually inductive, b is negative, and for narrow-band filters when b is numerically large, the electrical lengths of the cavities are chosen to be slightly less than π, i.e. a half-wavelength.

In order to obtain a simple design procedure for filters having two or more cavities, the principle will be adopted that any pair of adjacent cavities should be critically coupled. This is true, as will be seen, when the overall characteristic is of the maximally flat type (all roots of the insertion-loss function coincide) or of the Chebyshev type (all cavities identical), and will be retained as a design principle for more general types. It must be emphasized that the various overall filter characteristics are conveniently obtained by varying the relationships between the Q-factors of the individual cavities and not by a critical variation of the coupling lengths.

The conditions (16) are not sufficient for the required condition of critical coupling between the cavities. For this, $(L-1)$ must be quartic in first-order terms, i.e. in the deviation Δf from the mid-band frequency. It is readily proved by small-order variation that the expression $(m_1'n_2' - m_2'n_1')$ is quadratic in first-order terms for small deviations from the resonant conditions, but the last term in eqn. (15) requires the additional conditions $l_1' = l_2' = 0$, which, when combined with condition (16), give at the mid-band frequency

$$\tan\phi_i = -\frac{b_i}{2}, \; i = 1, 2 \quad . \quad . \quad . \quad . \quad (17)$$

where $0 < \phi_i < \pi/2$ is chosen for b_i negative. In the usual case of narrow-band filters the critical coupling length corresponding to $\tfrac{1}{2}(\phi_1 + \phi_2)$ is nearly equal to a quarter-wavelength, and at the resonant frequency,

$$\phi_i = \theta_i - \frac{\pi}{2} \quad . \quad . \quad . \quad . \quad (18)$$

(7) EQUIVALENCE BETWEEN DIRECT AND QUARTER-WAVE COUPLINGS

A quarter-wave-coupled filter may be transformed into a direct-coupled version by replacing the adjacent susceptances of each cavity and the quarter-wave coupling by a single large susceptance. Only the two susceptances at the extreme ends of the original filter remain identical. Any two networks which have equal insertion losses and phase shifts at a single frequency behave very similarly at that frequency (in fact, identically, if they are symmetric networks). The single large susceptance is chosen to give the same insertion loss as the quarter-wave coupling section, so that adjacent cavities practically remain critically coupled (exactly so, if the cavities are equal). Phase-

shift differences are compensated by increasing the lengths of the cavities in the direct-coupled filter. This means that the two filters possess slightly different characteristics, and since the direct-coupled filter cavities are longer, this filter will have a greater overall loaded Q-factor than the associated quarter-wave-coupled filter.
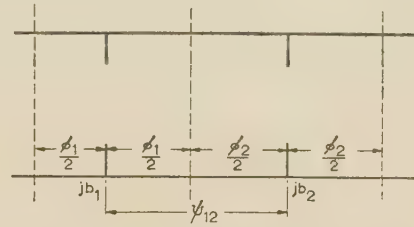
Fig. 4.—Quarter-wave coupling section.

Fig. 4 shows a quarter-wave coupling section with extra lengths of line $\phi_i/2$ on either side, given by eqn. (17). The transfer matrix for the susceptance jb_1 at the centre of a line of length ϕ_1 is

$$\begin{bmatrix} \cos\frac{\phi_1}{2} & j\sin\frac{\phi_1}{2} \\ j\sin\frac{\phi_1}{2} & \cos\frac{\phi_1}{2} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ jb_1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \cos\frac{\phi_1}{2} & j\sin\frac{\phi_1}{2} \\ j\sin\frac{\phi_1}{2} & \cos\frac{\phi_1}{2} \end{bmatrix} = \begin{bmatrix} A_1 & jB_1 \\ jC_1 & A_1 \end{bmatrix} \quad (19)$$

where
$$A_1 = \cos\phi_1 - \tfrac{1}{2}b_1\sin\phi_1$$
$$B_1 = \sin\phi_1 + \tfrac{1}{2}b_1(\cos\phi_1 - 1)$$
$$C_1 = \sin\phi_1 + \tfrac{1}{2}b_1(\cos\phi_1 + 1)$$

The phase shift at resonance is given by eqn. (7), and is equal to arc tan ϕ, i.e. π in this case. This result may be verified by adding the phase shift due to jb_1 alone to that given by ϕ_1,

i.e. $$\text{arc tan}\frac{b_1}{2} + \text{arc tan}\left(-\frac{b_1}{2}\right) = \pi$$

(where the function arc tan will only apply to angles between 0 and π).

The transfer matrix for the whole coupling section is

$$\begin{bmatrix} A_1 & jB_1 \\ jC_1 & A_1 \end{bmatrix}\begin{bmatrix} A_2 & jB_2 \\ jC_2 & A_2 \end{bmatrix} = \begin{bmatrix} A_1A_2 - B_1C_2 & j(A_1B_2 + B_1A_2) \\ j(A_1C_2 + C_1A_2) & A_1A_2 - C_1B_2 \end{bmatrix}$$
$$. \quad . \quad . \quad . \quad (20)$$

and hence the insertion loss is given by an expression of the form of eqn. (15), which reduces at resonance to

$$L = 1 + \tfrac{1}{4}[\tfrac{1}{2}b_1\sqrt{(b_2^2 + 4)} + \tfrac{1}{2}b_2\sqrt{(b_1^2 + 4)}]^2 \quad . \quad (21)$$

This is the same as the insertion loss of a single large susceptance jb' (b' usually negative), where

$$b' = \tfrac{1}{2}b_1\sqrt{(b_2^2 + 4)} + \tfrac{1}{2}b_2\sqrt{(b_1^2 + 4)} \quad . \quad . \quad (22)$$

The phase shift of the section shown in Fig. 4 is equal to 2π, so that the phase shift of the quarter-wave coupling section alone is

$$\psi_{12} = 2\pi - \tfrac{1}{2}(\phi_1 + \phi_2)$$
$$= \pi + \tfrac{1}{2}\left(\text{arc tan}\frac{b_1}{2} + \text{arc tan}\frac{b_2}{2}\right) \quad . \quad . \quad (23)$$

15

The phase shift of the single susceptance jb' is

$$\psi' = \arctan \frac{b'}{2} - \pi - \phi' \quad . \quad . \quad . \quad . \quad (24)$$

where $\phi' = \arctan(-b/2)$, and it is necessary to compensate for the phase difference $(\psi_{12} - \psi')$. This is achieved by replacing the whole section shown in Fig. 4 by the single susceptance jb' with coupling lines of length $\phi'/2$ on either side, as shown in Fig. 5. The two configurations then have the same *effective* phase shift at the resonant frequency, namely 2π and π, respectively.
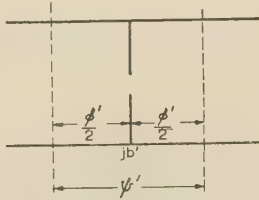
Fig. 5.—Direct-coupling section.
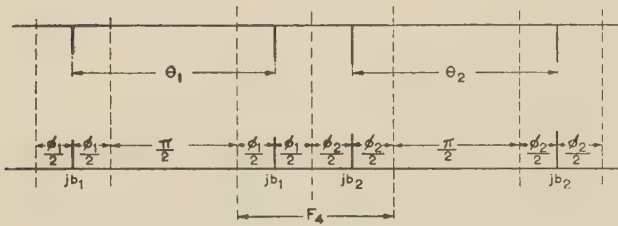
Fig. 6.—Section of a quarter-wave coupled filter, showing line lengths at mid-band.

Fig. 6 shows part of a quarter-wave-coupled filter (as in Fig. 3) and gives the line lengths at resonance. The resonant lengths of the cavities are given by eqns. (16), which may be rewritten in the form shown in the diagram, since

$$\arctan 2/b = \arctan(-b/2) + \pi/2$$
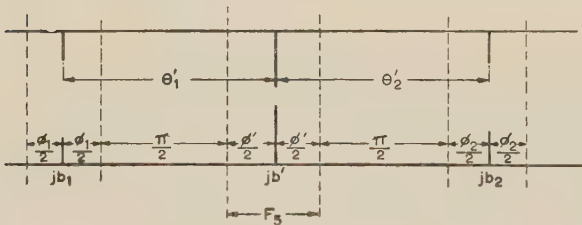
Fig. 7.—Section of a direct-coupled filter, showing line lengths at mid-band.

The part marked F_4 is replaced by that marked F_5 in Fig. 7 to form the associated direct-coupled filter. The lengths of the cavities here are

$$\theta_1' = \frac{\pi}{2} + \frac{\phi_1 + \phi'}{2} = \frac{1}{2}\left(\arctan\frac{2}{b_1} + \arctan\frac{2}{b'}\right) \quad (25)$$

and

$$\theta_2' = \frac{\pi}{2} + \frac{\phi_2 + \phi'}{2} = \frac{1}{2}\left(\arctan\frac{2}{b_2} + \arctan\frac{2}{b'}\right) \quad (26)$$

where $\arctan 2/b'$ (like $\arctan 2/b_1$ and $\arctan 2/b_2$) is usually between $\pi/2$ and π, and nearly equal to π.

This process may clearly be repeated; at every stage the resulting electrical length of each cavity is the mean of the two values of $\arctan 2/b$ for its bounding susceptances.

(5) THE Q-FACTOR OF A DIRECT-COUPLED FILTER

The results obtained in Section 4, e.g. eqns. (25) and (26) enable a relationship to be derived between the Q-factors of the direct-coupled and the associated quarter-wave-coupled filter. In the case of narrow-band filters, the Q-factor of a single cavity is easily shown to be equal to

$$Q = \frac{f_0}{2\Delta f} = \left(\frac{\lambda g_0}{\lambda_0}\right)^2 \frac{\theta_0}{2\Delta\theta} \quad . \quad . \quad . \quad . \quad (27)$$

where the subscript 0 refers to values at resonance, and the semi-increments are taken to the 3 dB transmission points. Thus the Q-factor is proportional to the resonant electrical length of the cavity for a given bandwidth $\Delta\theta$—a result which is to be expected intuitively.

For a given type of filter, consisting of an odd number of cavities, the total Q-factor is proportional to that of the central cavity, which for the direct-coupled filter is proportional to $\theta_0' = \arctan 2/b'$, where jb' is the value of its bounding susceptances. Thus the Q-factor of the direct-coupled filter is greater than that of the associated quarter-wave-coupled filter by the factor

$$\frac{Q_{DC}}{Q_{\lambda g/4}} = \frac{\arctan\dfrac{2}{b'}}{\arctan\dfrac{2}{b}} \quad . \quad . \quad . \quad . \quad (28)$$

where jb is the value of the susceptances of the central cavity in the associated quarter-wave-coupled filter.

Similarly, for an even number of cavities, the total Q-factor of the filter depends on the resonant value of θ for the central pair of cavities (which are made equal),

i.e.

$$\frac{Q_{DC}}{Q_{\lambda g/4}} = \frac{\dfrac{1}{2}\left(\arctan\dfrac{2}{b_1'} + \arctan\dfrac{2}{b_2'}\right)}{\arctan\dfrac{2}{b}} \quad . \quad . \quad (29)$$

where b_1' and b_2' are the susceptances bounding one of the central cavities of the direct-coupled filter.

In either case for narrow-band filters, for example when $Q \simeq 200$, i.e. $b = -10$, $b' = -102$ [see eqns. (65) and (60) of Appendix 10.1], the numerators in either eqn. (28) or (29) will be nearly equal to π for three or more cavities (actually $178 \cdot 8$ for $b' = -102$). The denominator will be rather less than π, actually $168 \cdot 7°$ in the example quoted, so that the above factors are equal to $178 \cdot 8/168 \cdot 7 = 1 \cdot 065$, i.e. the bandwidth of the direct-coupled filter is $6 \cdot 5\%$ less than that of the associated quarter-wave-coupled filter. This factor is approximately halved for two cavity filters since $b_1' = b$ in eqn. (29) for this case.

A complete proof of the formula for the Q-factor of a direct-coupled filter is given for the special case of two cavities in Appendix 10.1. The decrease in the bandwidth upon converting a quarter-wave-coupled filter into the direct-coupled form was first noted by Riblet,[2] and experimental confirmation for two and three-cavity filters is presented in Section 7.

(6) GENERALIZED MULTIPLE-CAVITY MICROWAVE FILTERS

(6.1) Introduction

The theory of microwave filters may be approached in one of two ways—the first directly by transmission-line theory, and the second by synthesis from conventional low-frequency circuits whose properties are known. The latter method has been

eveloped extensively, and methods are available for designing
lters with specified characteristics.[5] However, Riblet[2] has
ointed out the limitations of this approach for the design of
icrowave filters. It is desirable to use a direct transmission-
ne theory if possible in order to avoid the dangers inherent
in drawing analogies between conventional low-frequency
ircuits and microwave elements, and to avoid mathematical
omplications. In addition, the transmission-line approach is
ore satisfactory from the point of view of most microwave
ngineers. In practice, the direct method cannot be applied
asily to more than two cavities (e.g. as in Appendix 10.1), and a
ompromise has been adopted. It can be shown that a single
avity behaves similarly to a single lumped element of specified
equency characteristics, and transmission-line theory may be
pplied directly to networks consisting of cascaded lumped
ements representing quarter-wave-coupled microwave filters.

The multiple-cavity microwave filters which have been con-
dered previously in most papers are of two types. In the
rst, a number of identical cavities are coupled to give a filter
ith high rejection outside the pass band, but with dips or
ipples inside the band whose magnitude is an increasing function
f the number of filters. This is referred to as the Chebyshev
lter. The insertion loss of this type of filter is given by

$$L = 1 + [xU_n(x)]^2, \quad x = 2Q_{L(1)}\frac{\Delta f}{f_0} \quad . \quad . \quad . \quad (30)$$

here $U_{n(x)}$ is the rationalized Chebyshev polynomial of the
cond kind* and of degree $(n - 1)$, n is the number of
avities,[3] and $Q_{L(1)}$ is the loaded Q-factor of each cavity. The
verall loaded Q-factor of the filter is greater than $Q_{L(1)}$ by a
ctor which is readily calculated for any value of n. The
agnitude of the dips within the pass band is equal to 0·16 dB
r $n = 3$, and to nearly 1 dB for $n = 4$, so that the Chebyshev
lter is impracticable for more than three cavities in most appli-
tions. The insertion loss far from mid-band (referred to as
e *rejected band*) is given by

$$\mathscr{L}_R \equiv 10 \log_{10} L_R = [6(n - 1) + 20n \log_{10} x] \text{ decibels} \quad (31)$$

The second type of filter is the maximally flat type, designed
 eliminate the ripples within the pass band by ensuring coin-
idence of all roots of the insertion loss as a function of frequency.
he loaded Q-factors of individual cavities are tapered according
 the law

$$Q_{L(r)} = Q_T \sin\frac{(2r - 1)\pi}{2n}, \quad r = 1, 2, \ldots n \quad . \quad . \quad (32)$$

here $Q_{L(r)}$ is the loaded Q-factor of the rth cavity, and Q_T is
e overall loaded Q-factor of the filter. The insertion loss of
e maximally flat filter is

$$L = 1 + x'^{2n}, \quad x' = 2Q_T\frac{\Delta f}{f_0} \quad . \quad . \quad . \quad (33)$$

nd the insertion loss in the rejection band is less than that of the
rresponding Chebyshev filter with the same bandwidth.

The two types of filters may be considered as special cases
 a generalized symmetrical filter where the Q-factors of the
dividual cavities are related arbitrarily, e.g. for six cavities the
-factors take the values $Q_x, Q_y, Q_z, Q_z, Q_y, Q_x$, respectively.
 specifying an allowable pass-band tolerance, a filter with a
arper cut-off than the maximally flat type can be obtained by

* $U_{n(x)} = \dfrac{\sin n\beta}{\sin \beta}$, for $x = \cos \beta$

 $\dfrac{\sinh n\alpha}{\sinh \alpha}$, for $x = \cosh \alpha$

For $x \gg 1$, $U_{n(x)} \simeq (2x)^{n-1}$.

using the following theory, which is more simple and direct
than previous theories which use the low-frequency analogue
method.

(6.2) Three-Cavity Generalized Filters

The transfer matrix for a single filter stage consisting of two
susceptances jb separated by an electrical distance θ is given by
eqn. (10), and by application of eqn. (5) the complex voltage
insertion-loss ratio is found to be

$$R = (\cos\theta - b\sin\theta) + j[b\cos\theta + (1 - \tfrac{1}{2}b^2)\sin\theta] \quad (34)$$

Hence the insertion loss [eqn. (11)] is given by

$$L = 1 + x^2 \quad . \quad . \quad . \quad . \quad (35)$$

where

$$x = b(\cos\theta - \tfrac{1}{2}b\sin\theta) \quad . \quad . \quad . \quad (36)$$

and the insertion phase shift [eqn. (7)] is given by

$$\tan\psi = \frac{b\cos\theta + (1 - \tfrac{1}{2}b^2)\sin\theta}{\cos\theta - b\sin\theta} \quad . \quad . \quad (37)$$

At resonance, where $x = 0$, $\theta = \theta_0 = \arctan 2/b$, this reduces to

$$\tan\psi = -\frac{2}{b} = \tan(\pi - \theta_0) \quad . \quad . \quad (38)$$

A single normalized admittance $2jx$, where x is given by
eqn. (36), gives the same expression for the insertion loss at all
frequencies; and since $R = 1 + jx$, so that the insertion phase
shift is zero when $x = 0$, it has the same phase shift also if
placed in the centre of a line of electrical length given by eqn. (38)
at the resonant frequency only.

The overall transfer matrix of a generalized quarter-wave-
coupled filter is obtained by multiplication of transfer matrices,
each representing the characteristics of a component single
cavity with the correct critical coupling 'lengths' $\tfrac{1}{2}(\theta_0 - \pi/2)$
on either side [see eqn. (18)]. Each cavity may be replaced by
the single admittance $2jx$ flanked by coupling lengths of $\tfrac{1}{2}(\pi - \theta_0)$
[see eqn. (38)], so that the filter is now represented by a cascade
of admittances $2jx$ flanked by coupling lengths equal to

$$\frac{1}{2}\left(\theta_0 - \frac{\pi}{2}\right) + \frac{1}{2}(\pi - \theta_0) = \frac{\pi}{4}$$

each section of the cascade being represented by the transfer
matrix

$$[x] = \begin{bmatrix} -x & j(1 - x) \\ j(1 + x) & -x \end{bmatrix} \quad . \quad . \quad . \quad (39)$$

[see eqn. (44) of Reference 3].

The error introduced by the use of this transfer matrix is
only that due to the variation of the coupling lengths with
frequency, which can be shown to be very small, even in the
case of comparatively wide-band filters.

By expanding x as a function of θ in a Taylor series it can be
shown that, for small variations of θ about the resonant value
when terms of the order $(\Delta\theta)^3$ may be neglected, x is proportional
to $\Delta\theta$, and hence when terms of the order $(\Delta f)^2$ may be neglected
(which is true for narrow-band filters),

$$x = 2Q_x\frac{\Delta f}{f_0} \quad . \quad . \quad . \quad . \quad (40)$$

where Q_x is the loaded Q-factor of the cavity [since from
eqn. (35) $x = 1$ at the 3 dB points]. The representation of the
three-cavity filter is shown in Fig. 8. The loaded Q-factors of
the cavities are Q_x, Q_y, Q_x, where

$$x = 2Q_x\frac{\Delta f}{f_0}, \quad y = 2Q_y\frac{\Delta f}{f_0} \quad . \quad . \quad (41)$$
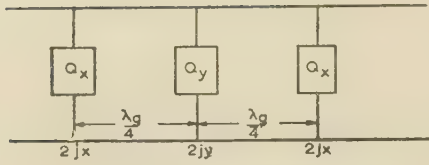
Fig. 8.—Three-cavity generalized filter.

The overall transfer matrix is

$$[x] \ [y] \ [x] =$$
$$\begin{bmatrix} -4x^2y + 2x + y & j(-4x^2y + 4xy - y + 2x - 1) \\ j(4x^2y + 4xy + y - 2x - 1) & -4x^2y + 2x + y \end{bmatrix}$$
$$. \quad . \quad . \quad . \quad (42)$$

Hence the insertion loss of the filter (eqn. 9) is given by

$$L = 1 + (4x^2y - 2x + y)^2 \quad . \quad . \quad . \quad (43)$$

or putting

$$\frac{Q_x}{Q_y} = k \quad . \quad . \quad . \quad . \quad . \quad (44)$$

$$L = 1 + (4k^2y^3 - 2ky + y)^2$$
$$= 1 + 16k^4y^2\left(y^2 - \frac{2k-1}{4k^2}\right)^2 . \quad . \quad . \quad (45)$$

This expression is valid only for narrow-band filters, when each cavity may be replaced by a susceptance $2jx$ or $2jy$. The bandwidth is given by the solutions of the equations

$$4k^2y^3 - 2ky + y = \pm 1 \quad . \quad . \quad . \quad . \quad (46)$$

Since the filter response is symmetrical for narrow-band filters with respect to the frequency variable y, it is valid to take only the positive sign and to double the value of Δf obtained, giving the bandwidth. The solution of eqn. (46) is given in
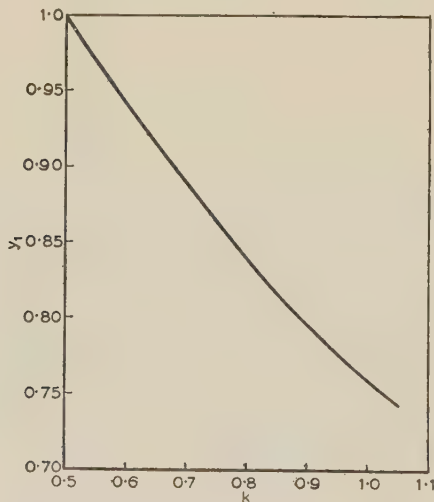
Fig. 9.—Solution of $4k^2y_1^3 - 2ky_1 + y_1 = 1$.

Fig. 9 for $k \geqslant \frac{1}{2}$. The value of $y = y_1$ corresponding to a given value of k is noted, and the overall Q-factor of the filter is given by

$$Q_T = Q_y/y_1 \quad . \quad . \quad . \quad . \quad . \quad (47)$$

since $2Q_T(\Delta f_T/f_0) = 1$, where $2Q_y(\Delta f_T/f_0) = y_1$.

Turning points in the characteristic are given by the solutions of the equation

$$\frac{d}{dy}(4k^2y^3 - 2ky + y)^2 = 0 \quad . \quad . \quad . \quad (48)$$

It is found that the dips in the pass-band transmission occur where $y^2 = (2k - 1)/12k^2$, and substitution of this value in eqn. (45) results in a pass-band tolerance (i.e. local maximum of loss in the pass band) of value

$$L_T = 1 + \frac{(2k-1)^3}{27k^2}, \text{ for } k \geqslant \frac{1}{2} \quad . \quad . \quad (49)$$

For large values of y, L may be approximated by retaining only the highest power of y,

i.e. $L_R = 16k^4y^6$, $\mathscr{L}_R = (60 \log_{10} y + 40 \log_{10} 2k)$ decibels (50)

where
$$y = 2Q_y\frac{\Delta f}{f_0} = 2y_1Q_T\frac{\Delta f}{f_0} = y_1y_T$$

Here y_1 is given by the solution of eqn. (46) and

$$y_T = 2Q_T\frac{\Delta f}{f_0} \quad . \quad . \quad . \quad . \quad (51)$$

Eqn. (50) may be written in the form

$$\mathscr{L}_R = [60 \log y_T + (60 \log y_1 + 40 \log 2k)] \text{ decibels} \quad (52)$$

which is useful in comparing the rejection-band characteristics of filters with different values of k, for the expression $(60 \log y_1 + 40 \log 2k)$ is independent of frequency—it depends only on k—and $60 \log y_T$ is a function of the frequency relative to the bandwidth only.

The Chebyshev and maximally flat filters may be seen as special cases of the general filter. The Chebyshev filter is the case $k = 1$ when

$$L = 1 + [y(4y^2 - 1)]^2 = 1 + [yU_3(y)]^2$$
$$y_1 = 0 \cdot 76$$

and $\mathscr{L}_r = 0 \cdot 16$ decibel

The maximally flat filter corresponds to $k = \frac{1}{2}$ when

$$L = 1 + y_T^6, \ y_1 = 1, \ \mathscr{L}_r = 0$$

and the attenuation in the rejection band is less than that of the corresponding Chebyshev filter by a factor of $4 \cdot 9$ dB.

The first step in constructing a filter is to decide on the maximum allowable pass-band tolerance, $\mathscr{L}_T$, thus fixing the maximum value of k. For example, in the case of the three-cavity filter it might be decided that a tolerance of $0 \cdot 04$ dB was permissible, giving $k = 0 \cdot 79$, so that the increase of attenuation in the rejection band would be $3 \cdot 5$ dB. As k increases beyond $0 \cdot 5$, $\mathscr{L}_T$ increases slowly from its zero value while $\mathscr{L}_R$ increases rapidly. Hence a considerable gain in rejection is possible at the expense of a very small increase in the pass-band tolerance.

The above theory has been extended to a larger number of cavities and is presented in Appendix 10.2. The design equations become increasingly complex, and, in fact, for five and six cavities they are functions of two independent variables. Filters of more than six cavities are rarely required.

(7) EXPERIMENTAL RESULTS

A number of two- and three-cavity direct-coupled filters have been constructed in order to test the two theories given in the paper, namely the expressions for the Q-factor of a direct-coupled filter and the insertion loss of a generalized filter. The

construction of the direct-coupled filters has been facilitated by the development of multiple-post structures.[6,7] Using five or more symmetrically placed posts across the guide, it is possible to obtain very high susceptances with sufficient accuracy.

For the measured values of the constructed susceptances of the direct-coupled filter in the sequence b_1, b_1' ... (from the outside of the filter towards the middle), the susceptances b_1, b_2, ... of the associated quarter-wave coupled filter may be derived by successive applications of eqn. (22). For this purpose it is more convenient to write eqn. (22) in the form

$$b' \simeq b_1 b_2 \left(1 + \frac{1}{b_1^2} + \frac{1}{b_2^2} \right); \quad b_1^2, b_2^2 \gg 1$$

The overall loaded Q-factor of the associated quarter-wave-coupled filter is calculated as described in the previous Section, and corrections are applied for the direct coupling [eqn. (28) or (29)] and the finite mid-band insertion loss. The latter factor is given approximately by[3]

$$\frac{Q \text{ (corrected for loss)}}{Q \text{ (calculated)}} = \frac{1}{\text{antilog } \frac{\alpha}{20}}$$

where α is the measured mid-band insertion loss in decibels.

It is difficult to check the theory of direct coupling with only two cavities, since the correction factor is only 3% for Q-factors of the order of 200, compared with about 6·5% for three or more cavities. The errors occur in the manufacture and measurement of the filter susceptances rather than in the measurement of bandwidth. However, a two-cavity filter has been designed accurately in accordance with the theory and gives rather better agreement with eqn. (64) than with eqn. (65) in Appendix 10.1. The results for the filter readjusted to operate at two frequencies are shown in Table 1.

Table 1

Outer susceptances b_1	Inner susceptance b'	Theoretical $Q_{\lambda g/4}$	Mid-band insertion loss	Corrected $Q_{\lambda g/4}$	Corrected Q_{DC}	Measured Q_{DC}
			dB			
− 9·65	− 95	174	0·4	166	171	170
−10·2	−104·5	210·8	0·7	195·6	201	198

Better confirmation has been obtained with three-cavity filters, several of which have been designed and measured, giving the five results shown in Table 2.

(8) CONCLUSIONS

A direct-coupled filter is generally preferable to the equivalent quarter-wave-coupled form because it is shorter and possesses lower mid-band insertion loss. A design for any specification is quickly established using a theory which is simpler and more exact than existing ones. In addition, an improvement over the usual maximally flat type of characteristic is possible by making use of a generalized theory of multiple-cavity filters. Good agreement between theory and experiment may be obtained.

(9) REFERENCES

(1) HESSEL, J., GOUBAU, G., and BATTERSBY, L. R.: 'Microwave Filter Theory and Design', *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 990.
(2) RIBLET, H. J.: 'Synthesis of Narrow-Band Direct-Coupled Filters', *ibid.*, 1952, **40**, p. 1219.
(3) PRITCHARD, W. L.: 'Quarter-Wave Coupled Waveguide Filters', *Journal of Applied Physics*, 1947, **18**, p. 862.
(4) MUMFORD, W. W.: 'Maximally-Flat Filters in Waveguide', *Bell System Technical Journal*, 1948, **27**, p. 684.
(5) RAGAN, G. L.: 'Microwave Transmission Circuits' (McGraw-Hill, 1948), p. 699.
(6) SIMON, J. C., and BROUSSAUD, G.: 'Les filtres passe-bande en hyperfréquence', *Annales de Radioélectricité*, 1953, **8**, p. 3.
(7) CRAVEN, G., and LEWIN, L.: 'Design of Microwave Filters with Quarter-Wave Couplings', *Proceedings I.E.E.*, Paper No. 2001 R, March, 1956 (**103** B, p. 173). Discussion, *ibid.* (**103** B, p. 632).

(10) APPENDICES

(10.1) Q-Factor of a Two-Cavity Direct-Coupled Filter

The filter consists of normalized susceptances jb, jb', and jb separated by electrical distances θ, and the transfer matrix for this configuration is

$$\begin{bmatrix} 1 & 0 \\ jb & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & j \sin \theta \\ j \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jb' & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & j \sin \theta \\ j \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jb & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ jb & 1 \end{bmatrix} \begin{bmatrix} A & jB \\ jC & A \end{bmatrix} \begin{bmatrix} 1 & 0 \\ jb & 1 \end{bmatrix}$$

where

$$A = \cos^2 \theta - \sin^2 \theta - b' \cos \theta \sin \theta$$
$$B = 2 \cos \theta \sin \theta - b' \sin^2 \theta$$
$$C = 2 \cos \theta \sin \theta + b' \cos^2 \theta$$

Table 2

b_1	b' (measured values)	b_2 eqn (22)	Q_x	Q_y	k	y_1	$\frac{Q_y}{y_1}$	Mid-band loss	Corrected $Q_{\lambda g/4}$	Corrected Q_{DC}	Measured Q_{DC}
								dB			
−8·8	− 95	−10·6	100·3	146·5	0·687	0·891	164	0·65	152	162	161
−9·4	−104·5	−10·9	126·2	170	0·745	0·876	194	0·95	169	179·5	180
−9·1	−106·5	−11·5	108	174	0·622	0·93	187	0·82	170	181	182
−9·65	−115	−10·9	134	170	0·79	0·845	200·5	1·1	180·5	192	207
−9·4	−95	−9·32	116	114	1·018	0·753	150·5	0·7	139	149	148

In every case the measured value is in much better agreement with Q_{DC} than with $Q_{\lambda g/4}$. This lends confirmation both to the theory of direct coupling and to that of generalized filters. The finite mid-band insertion loss tends to mask the theoretical pass-band tolerance $\mathscr{L}_T$, but very good agreement between theory and experiment, usually to better than 1 dB, has been obtained for the attenuation in the rejection band, $\mathscr{L}_R$.

Thus the final transfer matrix is

$$\begin{bmatrix} A' & jB' \\ jC' & A' \end{bmatrix}$$

where

$$A' = A - bB$$
$$B' = B$$
$$C' = 2bA - b^2 B + C$$

The insertion loss of the filter, given by eqn. (9), is

$$L = 1 + \tfrac{1}{4}(C' - B')^2$$

where
$$C' - B' = 2bA - b^2B + (C - B)$$
$$= 2b(\cos^2 \theta - \sin^2 \theta - b' \cos \theta \sin \theta)$$
$$- b^2(2 \cos \theta \sin \theta - b' \sin^2 \theta) + b' . \quad (53)$$

In order to give critical coupling the equation $C' - B' = 0$, which is quadratic in $\tan \theta$, must have equal roots, i.e. the following equations must be satisfied:

$$C' - B' = 0 \quad . \quad . \quad . \quad . \quad . \quad (54)$$

$$\frac{d}{d\theta}(C' - B') = 0 \quad . \quad . \quad . \quad . \quad (55)$$

It is more convenient to express $(C' - B')$ in terms of the double angle;

i.e. $C' - B' = \dfrac{b}{2}(4 - bb') \cos 2\theta$
$$- b(b + b') \sin 2\theta + \tfrac{1}{2}b'(b^2 + 2) \quad (56)$$

so that eqn. (55) becomes

$$(4 - bb') \sin 2\theta + 2(b + b') \cos 2\theta = 0 \quad . \quad . \quad (57)$$

i.e. $\tan 2\theta = \dfrac{\dfrac{2}{b} + \dfrac{2}{b'}}{1 - \dfrac{2}{b}\dfrac{2}{b'}}$

or $\theta = \dfrac{1}{2}\left(\arctan\dfrac{2}{b} + \arctan\dfrac{2}{b'}\right) \quad . \quad . \quad . \quad (58)$

Substitution of $(b + b')$ from eqn. (57) into eqn. (58) gives

$$C' - B' = -\frac{b}{2}(bb' - 4)(\cos 2\theta + \sin 2\theta \tan 2\theta) - \frac{b'}{2}(b^2 + 2)$$

$$= -\frac{b}{2}(bb' - 4)\sec 2\theta - \frac{b'}{2}(b^2 + 2) \quad . \quad . \quad . \quad (59)$$

Hence, from eqns. (57) and (59), and using $\sec^2 2\theta = 1 + \tan^2 2\theta$, eqn. (54) becomes

$$\frac{b'^2(b^2 + 2)^2}{b^2} = (bb' - 4)^2 + 4(b + b')^2$$

i.e. $b'^2(b^2 + 2)^2 = b^2(b^2 + 4)(b'^2 + 4)$

which finally reduces to the form of eqn. (22);

i.e. $b' = b\sqrt{(b^2 + 4)} \quad . \quad . \quad . \quad . \quad (60)$

The 3 dB points occur when

$$L = 1 + \tfrac{1}{4}(C' - B')^2 = 2$$

i.e. when $|C' - B'| = 2$. The Q-factor of the filter is given by eqn. (27), where $\Delta\theta$ is found by expanding $(C' - B')$ in a Taylor series, i.e. writing

$$(C' - B') = f(\theta)$$

$$f(\theta) = f(\theta_0) + f'(\theta_0)\Delta\theta + f''(\theta_0)\frac{(\Delta\theta)^2}{2} + 0(\overline{\Delta\theta}^3)$$

where θ_0 is the resonant value of θ given by eqn. (58). Since $f(\theta_0) = f'(\theta_0) = 0$, $\Delta\theta$ is given by the expression

$$|f''(\theta_0)|\frac{(\Delta\theta)^2}{2} = 2 \quad . \quad . \quad . \quad . \quad (61)$$

Thus, from eqns. (27) and (61),

$$Q = \frac{1}{4}\left(\frac{\lambda g_0}{\lambda_0}\right)^2 \theta_0 \sqrt{|f''(\theta_0)|} \quad . \quad . \quad . \quad . \quad (62)$$

Differentiation of eqn. (56) gives

$$-\frac{1}{b}f'(\theta) = (4 - bb') \sin 2\theta + 2(b + b') \cos 2\theta$$

and $$-\frac{1}{2b}f''(\theta) = (4 - bb') \cos 2\theta - 2(b + b') \sin 2\theta.$$

Putting $\theta = \theta_0$, using $f'(\theta_0) = 0$, and eliminating b_0, it is found that

$$|f''(\theta_0)| = 2|b|\sqrt{[(bb' - 4)^2 + 4(b + b')]}$$
$$= 2|b|\sqrt{[(b^2 + 4)(b'^2 + 4)]}$$

i.e. $|f''(\theta_0)| = 2|b'|\sqrt{[b'^2 + 4]}$

For narrow-band filters $|b'|$ is large, and using eqn. (60) this reduces to

$$|f''(\theta_0)| \simeq 2b'^2 = 2(b^4 + 4b^2) \quad . \quad . \quad (63)$$

where this approximation has a proportional error of only $2/b^4 \times 100\%$, which is very small.

Substitution of θ_0 (eqn. 58) and $|f''(\theta_0)|$ in eqn. (62) gives the final expression for Q, i.e.

$$Q_{DC} = \frac{1}{4}\sqrt{2}\left(\frac{\lambda g_0}{\lambda_0}\right)^2 \sqrt{(b^4 + 4b^2)}\frac{1}{2}\left(\arctan\frac{2}{b} + \arctan\frac{2}{b'}\right)$$
$$\quad . \quad . \quad . \quad (64)$$

Eqn. (64) is accurate to within a proportional error of $2/3b^4 \times 100\%$, which is negligible for narrow-band filters. This can be seen by expanding $f(\theta)$ to the fifth term, as $f'''(\theta_0)$ vanishes.

The Q-factor for the associated quarter-wave-coupled filter is readily shown to be

$$Q_{\lambda g/4} = \frac{1}{4}\sqrt{2}\left(\frac{\lambda g_0}{\lambda_0}\right)^2 \sqrt{(b^4 + 4b^2)} \arctan\frac{2}{b} \quad . \quad (65)$$

and hence Q_{DC} is greater than $Q_{\lambda g/4}$ by the factor

$$\frac{\dfrac{1}{2}\left(\arctan\dfrac{2}{b} + \arctan\dfrac{2}{b'}\right)}{\arctan\dfrac{2}{b}} \quad . \quad . \quad . \quad . \quad (66)$$

which is analogous to eqn. (29).

(10.2) **Generalized Theory for Four, Five and Six Cavities**

(10.2.1) **Four Cavities.**

The transfer matrix is

$$[x] \ [y] \ [y] \ [x]$$

$$= [x]\begin{bmatrix} A_1 & jB_1 \\ jC_1 & A_1 \end{bmatrix}[x]$$

$$= \begin{bmatrix} A_2 & jB_2 \\ jC_2 & A_2 \end{bmatrix}$$

where $A_1 = 2y^2 - 1, B_1 = 2y(y - 1), C_1 = -2y(y + 1)$
$$A_2 = (2A_1 + B_1 - C_1)x^2 + (B_1 + C_1)x - A_1$$
$$B_2 = (2A_1 + B_1 - C_1)x^2 - 2(A_1 - C_1)x - C_1$$
$$C_2 = -(2A_1 + B_1 - C_1)x^2 - 2(A_1 + B_1)x - B_1.$$

Therefore $\quad B_2 - C_2 = 4(4y^2 - 1)x^2 - 8yx + 4y^2$

and the insertion loss is

$$L = 1 + 4[(4y^2 - 1)x^2 - 2yx + y^2]^2$$

or for $x = ky$, i.e. eqn. (44),

$$L = 1 + 4y^4[(4y^2 - 1)k^2 - (2k - 1)]^2 \quad . \quad . \quad (67)$$

The bandwidth of the filter is given by the solution of the equation

$$2y_1^2[(4y_1^2 - 1)k^2 - (2k - 1)] = 1 \quad . \quad . \quad . \quad (68)$$

which is given in graphical form in Fig. 10. The overall loaded Q-factor is

$$Q_T = Qy/y_1 \quad . \quad . \quad . \quad . \quad . \quad (69)$$
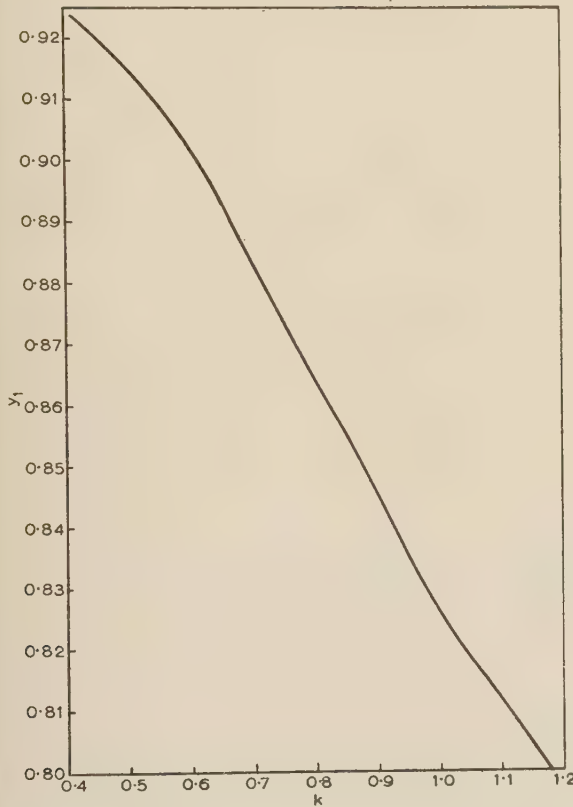
Fig. 10.—Solution of $2y_1^2[(4y_1^2 - 1)k^2 - (2k - 1)] = 1$.

The pass-band tolerance is

$$L_T = 1 + \frac{1}{64k^4}(k^2 + 2k - 1)^4 \quad . \quad . \quad . \quad (70)$$

and the attenuation in the rejection band is

$$\mathscr{L}_R = (80 \log_{10} y + 10 \log_{10} 2^6 k^4) \text{ decibels} \quad . \quad . \quad (71)$$

or putting $y = y_T y_1$, where $y_T = 2Q_T \dfrac{\Delta f}{f_0}$,

$$\mathscr{L}_R = [80 \log y_T + 40 \log 2\sqrt{(2)}\, ky_1^2] \text{ decibels} \quad . \quad (72)$$

The case of the maximally-flat filter is obtained by putting $\mathscr{L}_T = 0$, i.e. $k^2 + 2k - 1 = 0$, giving $k = \sqrt{(2)} - 1$, which when substituted in eqn. (68) gives

$$y_1^2 = \frac{1}{2\sqrt{2}[\sqrt{(2)} - 1]}$$

Hence from eqn. (72) the attenuation in the rejection band is $80 \log y_T$ decibels in agreement with the formula $L = 1 + y_T^8$ for the four-cavity maximally-flat filter. The difference in $\mathscr{L}_R$ for the Chebyshev and maximally flat filter is $11 \cdot 44$ dB.

(10.2.2) **Five Cavities.**

The theory given for three and four cavities may be applied to a larger number, but the design equations involve functions of several variables. The transfer matrix for five cavities is obtained by a simple extension of that for three cavities, and the insertion loss for a quarter-wave-coupled filter where the Q-factors take the values Q_x, Q_y, Q_z, Q_y, Q_x, is

$$L = 1 + [8x^2(2y^2z - y) - 2x(4yz - 1) + (4y^2z - 2y + z)]^2$$
$$. \quad . \quad . \quad (73)$$

Putting $\quad k = \dfrac{Q_y}{Q_z} = \dfrac{y}{z}, \quad l = \dfrac{Q_x}{Q_z} = \dfrac{x}{z} \quad . \quad . \quad . \quad (74)$

$$L = 1 + z^2[16\overline{lk}^2 z^4 - 4kXz^2 + Y]^2 \quad . \quad . \quad (75)$$

where $\quad X = 2l^2 + 2l - k \quad . \quad . \quad . \quad . \quad (76)$

$$Y = 2l - 2k + 1 \quad . \quad . \quad . \quad . \quad (77)$$

The overall loaded Q-factor is given by the solution of the equation

$$16(lk)^2 z_1^5 - 4kXz_1^3 + Yz_1 = 1 \quad . \quad . \quad . \quad (78)$$

For five cavities there are four ripples in the pass band, i.e. two unequal pairs, with insertion loss given by

$$L_T = 1 + \tfrac{16}{25}z^2(2kXz^2 - Y)^2$$

where $\quad z^2 = \dfrac{3X \pm \sqrt{(9X^2 - 20l^2Y)}}{40l^2k} \quad . \quad . \quad . \quad (79)$

and the attenuation in the rejection band is

$$\mathscr{L} = (100 \log z + 40 \log 4lk) \quad . \quad . \quad . \quad (80)$$

$$= [100 \log z_T + (100 \log z_1 + 40 \log 4lk)] \text{ decibels} \quad . \quad (81)$$

The maximally flat case is obtained by solving the equations $X = Y = 0$, obtaining

$$l = \frac{\sqrt{(5)} - 1}{4} = \sin 18°$$

$$k = \frac{\sqrt{(5)} + 1}{4} = \sin 54°$$

and $\quad L = 1 + z_T^{10}.$

(10.2.3) **Six Cavities.**

The loaded Q-factors are Q_x, Q_y, Q_z, Q_z, Q_y, Q_x, respectively. The transfer matrix is obtained by an extension of that for four cavities, which results in the following expression for the insertion loss:

$$L = 1 + 4z^4[16(kl)^2 z^4 - 4kX'Z^2 + Y']^2 \quad . \quad . \quad (82)$$

where k and l are defined in eqn. (74), and

$$X' = kl^2 + 2l^2 + 2l - k \quad . \quad . \quad . \quad (83)$$

$$Y' = l^2 + 2lk + 2l - k^2 - 2k + 1 \quad . \quad (84)$$

The bandwidth is obtained by a solution of the equation

$$2z_1^2[16(kl)^2 z^4 - 4kX'z^2 + Y'] = 1 \quad . \quad . \quad (85)$$

The magnitudes of the two pairs of ripples in the pass band are given by the expression

$$L_T = 1 + \tfrac{16}{9}z^4(2kX'z^2 - Y')^2$$

where

$$z^2 = \frac{X' \pm \sqrt{(X'^2 - 3l^2Y')}}{12l^2k} \right\} \quad . \quad . \quad . \quad (86)$$

and the attenuation in the rejection band is

$$\mathscr{L} = [120 \log z + 40 \log 4\sqrt{(2)kl}] \quad . \quad . \quad . \quad . \quad (87)$$

$$= [120 \log z_T + 40 \log 4\sqrt{(2)klz_1^3}] \text{ decibels} \quad . \quad . \quad (88)$$

The maximally flat filter is obtained by solving the equations $X' = Y' = 0$, obtaining

$$l = 2 - \sqrt{3} = \sin 15°/\sin 75°$$

$$k = \sqrt{(3)} - 1 = \sin 45°/\sin 75°$$

and

$$L = 1 + 4^3(kl)^4 z_T^{12} = 1 + z_T^{12}$$

since

$$z = z_T \sin 75° = z_T\sqrt{\frac{\sqrt{(3)} + 2}{4}}.$$

A GRAPHICAL APPROACH TO THE STUDY OF IRREGULARITIES IN TRANSMISSION LINES

By W. T. BLACKBAND, M.Sc., Associate Member.

SUMMARY

The effects of an irregularity of characteristic impedance in a transmission line are displayed graphically on a transmission-line circle diagram. This presentation shows qualitatively the major factors which determine the changes in voltage standing-wave ratio and phase produced by a single irregularity or by a large number of irregularities distributed along the line. This analysis is not limited to matched transmission lines as are the statistical methods. The resonance of a disc-spaced transmission line at super-high-frequency is considered. The method developed by Cotte for the determination of impedance variation along a transmission line by harmonic analysis of the distribution of irregularities is discussed in terms of the graphical approach, and a possible simplification to the experimental method is suggested.

(1) INTRODUCTION

There have been many studies made of the effects of irregularities of characteristic impedance in long transmission lines. Most of these have used statistical methods which lead to conclusions not easy to verify experimentally. The present approach is admittedly only qualitative, but it is presented here because it throws some light on the mechanism of reflection in transmission lines and shows the three fundamental factors governing the total reflected wave at the input.

For most studies of irregularities it is assumed that the transmission line is terminated in its mean characteristic impedance, i.e. that it would carry no standing wave were it not for the internal reflections. The present approach is more general. In this we study the change in standing-wave ratio produced by the irregularities in a line mismatched at its termination. Consider the length of line AB in Fig. 1(a): this is not matched at A, and consequently it supports a voltage standing-wave along its length. If the length BC were of the same characteristic impedance as AB and CD (i.e. $Z_0' = Z_0''$), then the AB standing-wave pattern would extend through BC and CD. However, the effect of the change in characteristic impedance over BC will be to disturb the standing-wave pattern in CD. The reflection coefficient at C may be increased or decreased in magnitude and may be changed in phase. These changes in the reflection coefficient are manifest in corresponding changes in the standing-wave pattern in CD. An increase in the magnitude of the reflection coefficient increases the v.s.w.r., and a change in phase moves the pattern along the transmission line. The movement of the pattern towards the load is a sign that the electrical length of the system is increased, and there is 'delay'; movement toward the generator shows that the electrical length is lessened and there is 'advance'. The standing-wave pattern thus suffices to define all effects of the irregularity of characteristic impedance.

The transmission-line circle diagram is a powerful tool in the representation and solution of problems connected with standing waves on transmission lines, and in the following Sections the effects of irregularities are discussed in terms of the circle diagram.

The properties of the Cartesian form of circle diagram have been described by Willis Jackson,[1] Huxley[2] and Slater.[3] Reference should be made to such literature for a detailed account.

Briefly, this form of circle diagram consists of a Cartesian grid for R and X on which two families of circles are drawn. These circles define constant values of u and v, which variables are
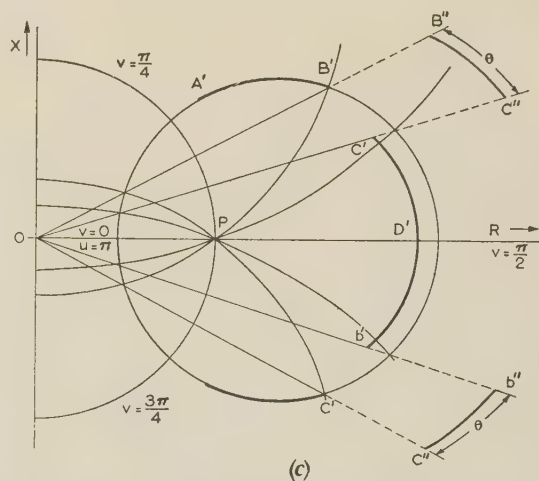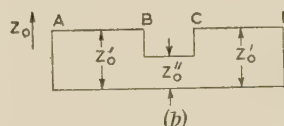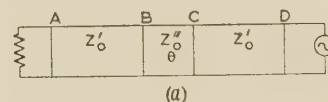
Fig. 1.—Single impedance irregularity.

(a) Transmission line with irregularity.
(b) Variation of Z_0 along line.
(c) V.S.W.R. diagram for the irregularity.

defined by $R + jX = \tanh(u + jv)$. The u-circles rest one within another about the unit point P. The v-circles all pass through P, cutting the u-circles orthogonally. Only the half of the grid for R positive is of practical importance. In Fig. 1(c) this half is divided about the point P, into four quadrants by the R-axis and the unit circle (this is the circle which corresponds to $v = \pi/4$ or $3\pi/4$). To any point $R + jX$ correspond values of u and v such that $R + jX = \tanh(u + jv)$. But $1/\tanh(u + jv) = \tanh u + (jv + \pi/2)$, and therefore the point $1/(R + jX)$ is represented by the intersection of the circles u and $v + \pi/2$. This means that the reciprocal of any point lies in the diagonally opposite

quadrant, and the reciprocal of the grid in any quadrant is that in the quadrant diagonally opposite. This reciprocal symmetry may be more evident upon reference to the polar form of the diagram,[4,5] in which the grids are transformed so that the R-axis and the loci of $v = \pi/4$ and $v = 3\pi/4$ are straight lines at right angles and pass through the mid-point.

The special property of the circle diagram which is used here is that if the impedance at points in a transmission line are plotted taking its characteristic impedance as a unit, then these impedances lie upon a circle of constant u, and the v-circles mark off a scale of electrical length along the transmission line, one circuit of the circle corresponding to an increase of v by π radians (i.e. a distance of $\lambda/2$). In moving along the line from load to signal source the u-circle is traversed in a clockwise direction.

In Sections 2 and 3 irregularities are considered which are of 'rectangular' form, i.e. they consist of a short length of line of which the characteristic impedance is the same throughout but differs from that of the rest of the line. Figs. 1(a) and 1(b) show such a 'rectangular' irregularity. The irregularities inadvertently introduced during the manufacture of a radio-frequency cable are not of this form. In general, such an irregularity decreases smoothly from its peak and merges imperceptibly into the undisturbed line, or into an adjacent irregularity. The statistical analyses made of the effects of irregularities are based upon two parameters of the irregularities, their amplitude and 'correlation range'. The latter is a measure of the length of the irregularity, but it has been defined differently by the various authors,[6-8] depending upon the law assumed for the decay of the irregularity on either side of its maximum value.

The assumption of a rectangular form is convenient for a graphical analysis, and any more complex form can be imagined as broken down into a series of rectangular steps. There is in this case no doubt as to the meaning of 'correlation range'—it is the length of the irregularity.

(2) EFFECT OF A SINGLE IRREGULARITY

Consider a length of line which has a uniform characteristic impedance except over a short length where there is an irregularity of impedance. For simplicity it is assumed that the irregularity consists of an abrupt change in characteristic impedance of 'rectangular' form. Such an irregularity is shown in Figs. 1(a) and 1(b) where the section BC of electrical length θ and characteristic impedance Z_0'' is in a line AD of otherwise uniform characteristic impedance, Z_0'. The effects of this irregularity are to change the v.s.w.r. in CD and to change the phase of the wave in CD compared with what it would be if BC were uniform in Z_0 with the rest of the line. These two effects are considered separately.

The effect on v.s.w.r. of this irregularity is shown in terms of the impedance circle diagram of Fig. 1(c). Suppose that the line is terminated in an impedance such that the impedances of points between A and B are represented by points in the arc A'B' where the plotted impedances are normalized with respect to Z_0', the characteristic impedance of AB. In order to represent graphically the impedance changes in the length of line BC by the arc of a u-circle it is necessary to plot impedances normalized with respect to Z_0'', the characteristic impedance of BC. The impedance represented by B' must now be represented by B'', where OB'B'' is a straight line and $OB'/OB'' = Z_0''/Z_0'$. The impedance changes in BC are represented by the arc B''C'', the length of which represents θ, the electrical length of BC. In order to interpret the impedance C'' in terms of the original characteristic impedance Z_0' the scaling process is reversed: $OC'/OC'' = Z_0''/Z_0'$. The changes in impedance beyond C are shown by the arc C'D'.

It will be noted that an overall effect is a change in the v.s.w.r. of the line. In this case the v.s.w.r. in CD has been reduced by the irregularity. This is shown graphically by the radius of C'D' being smaller than that of A'B'.

If a second similar irregularity is placed in the line, as represented by b'b''c''c' in Fig. 1(c), and if the two irregularities are equally spaced either side of the R-axis, it is evident from the symmetry of the diagram that the second irregularity will increase the v.s.w.r. by the same factor as the first decreased this ratio, and that such a pair of irregularities would have no net effect upon the v.s.w.r. in the line. If the two irregularities are brought closer together until they meet to form one irregularity with its centre on the R-axis, their net effect on the v.s.w.r. will still be zero, and it is thus evident that an irregularity on the R-axis will not affect the v.s.w.r. Reference to Fig. 2 shows that the
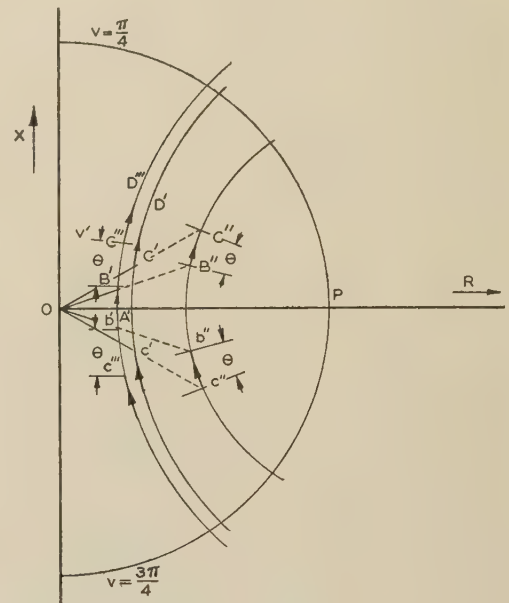
Fig. 2.—Phase diagram for an irregularity.

same symmetry holds at the low-impedance crossing of the circle and R-axis. From the above it can be seen that the change in v.s.w.r. produced by an irregularity is of different signs in the inductive and capacitive halves of the impedance plane, and is zero on the R-axis. Because of the reciprocal symmetry of the parts of the impedance plane inside and outside the unit circle (circle of $v = \pi/4$ or $3\pi/4$), the v.s.w.r. produced by an irregularity on the unit circle must be either a maximum or a minimum—as there is no reason to expect maxima anywhere else it is fair to assume that the maximum occurs on the unit circle.

The effects of the irregularity on the phase of the signal are considered with the help of Fig. 2, where the points A'B'B''C''C' and D' have the same significance as in Fig. 1(c), and C'''D''' is the arc which would have represented the impedance changes if BC were not irregular in impedance. On this assumption B'C''' would represent a phase length θ measured round the u-circle. As the v-circles, by which the phase length of the arcs are measured, pass through the point P, they must be closer together for small u-circles than for large ones. For this reason the arc B''C'' will be shorter than B'C'''. The arc B'C' will be further shortened from the length B''C'' in the process of renormalizing by a factor of Z_0''/Z_0'. In Fig. 2 the circle v' is marked. This passes through C''' and represents the phase at C if BC is regular. It is noticed that C' is below this arc, and that it is

necessary to move along some distance from C' towards D' before the v'-circle is crossed. This distance represents a phase delay accompanying transmission through BC, due to the reduction of Z_0 in that length. A study of the impedance locus c'c''b''b' in the adjacent quadrant of Fig. 2 shows that in this quadrant there is also a delay introduced in traversing the discontinuity.

If the irregularity is small in electrical length and equally disposed about the $v = \pi/4$ circle, it is easy to see that there would be no phase change because the lines B'B'' and C'C'' would be normal to that circle. If this irregularity were large it could be shown by reasoning similar to that in the previous paragraph that there would be a phase delay in traversing the part on the near side of the $v = \pi/4$ circle and a phase advance in traversing the far part of the irregularity. From this we may generalize that irregularities between the $v = 0$ and $\pi/4$ and $v = 3\pi/4$ and 0 circles give phase delay, while those between the $v = \pi/4$ and $3\pi/4$ circles give phase advance. It is also evident from the reciprocal equivalence between the parts of the impedance plane inside and outside the unit circle $v = \pi/4$ (or $3\pi/4$) that a given discontinuity will produce equal and opposite phase changes at points v and $v + \pi/2$.

It will be seen that a given irregularity will cause the greatest disturbance in phase when it is situated in a transmission line about a point of resistive impedance, and least at points $\pi/4$ (i.e. $\lambda/8$) away from the points of pure resistance. For the change in v.s.w.r. the effects are least at the points of pure resistance and greatest $\pi/4$ (i.e. $\lambda/8$) away. In each case the effects change sign in passing through their zero value.

The effect of the irregularity BC is to send a reflected wave from C towards D, and a second wave, of almost equal amplitude and changed almost $\pi/2$ in phase, from B towards D. The net effect will be a change in the v.s.w.r. in BC where there is only one reflection, and a partial cancellation of the two reflections in the region CD. The shorter BC, the nearer the two reflections are towards equality in modulus and phase, and the closer the cancellation in CD. The longer BC, the less cancellation, until as BC becomes $\lambda/4$ long ($\theta = \pi/2$) the irregularity has its maximum effect, because the phase delay between the waves from B and C, due to the double transit of BC by the B wave, brings the two into phase. Thus a very great ratio of Z_0'/Z_0'' may cause but a negligible reflection if θ is small. This can be seen from Fig. 1(c) and 2, where it is clear that when B''C'' is small so must be B'C' (and v.s.w.r. and phase effects) even if OB'/OB'' is great.

It can be seen that when the ratio Z_0'/Z_0'' is nearly unity, the irregularity has a small effect, but that for Z_0'/Z_0'' far from unity the disparity in the radii of the arcs A'B' and B''C'' is great and the irregularity has a greater effect. Thus we can conclude that the effect of the irregularity is dependent upon three factors:

(a) The amplitude of the irregularity, Z_0'/Z_0''.
(b) The electrical length of the irregularity, θ.
(c) The position of the irregularity relative to the standing-wave pattern looking towards the load.

The effect of the position of the irregularity may be summarized as follows:

From $v = 0$ to $\pi/2$ an increase in Z_0 increases the v.s.w.r., and from $v = \pi/2$ to π an increase in Z_0 decreases the v.s.w.r. Also, from $v = \pi/4$ to $3\pi/4$ an increase in Z_0 gives phase delay, and from $v = 0$ to $\pi/4$ (and $v = 3\pi/4$ to π) an increase in Z_0 gives phase advance.

(3) EFFECT OF MANY IRREGULARITIES

If there are many small irregularities distributed along the transmission line, each will contribute a disturbance to the standing-wave pattern. The overall disturbance will depend upon the distribution of the irregularities along the line.

Two distributions are of practical interest, the 'uniform' and the 'equidistant' distribution (which is that of the spacers arranged along a disc-spaced coaxial line).

The 'uniform' distribution must be defined further. The distribution is uniform for the present purposes when the irregularities are many and are evenly distributed with respect to electrical length around the impedance circle (i.e. with respect to the standing-wave pattern). With such a distribution the change in v.s.w.r. due to the irregularities in the positive-reactance half of the circle is balanced by the equal and opposite change in v.s.w.r. due to those in the negative-reactance half. Likewise the phase change due to irregularities within the unit circle will be balanced by the opposite change in phase due to those outside. (Any non-uniform distribution which has symmetry about both the R-axis and the unit circle will give this same balance between the effects of the irregularities in the opposite halves of the circle.)

In illustration of this, consider a section $\lambda/2$ long in a length of line and add in this section irregularities such as those considered in Section 2. When the irregularities are evenly distributed and tightly packed, the $\lambda/2$ section of line will be all of characteristic impedance Z_0''. Because of the uniformity of this distribution of irregularities these should not give any disturbance in the system outside the $\lambda/2$ length; this, of course, is so, as it is well known that a half-wavelength of line of any impedance does not disturb the standing-wave pattern in the rest of any line in which it be inserted.

In order to be 'uniformly' distributed in this special sense the irregularities need not be uniformly distributed along the physical length of the line. Consider a length of line which has such a 'uniform' distribution of irregularities. It would be possible to cut this line anywhere and insert a half-wavelength of regular line in the break without disturbing the 'uniformity' with respect to the standing-wave pattern—however, the irregularity-free half-wavelength would rob the line of its physical uniformity. Likewise, if the line is several half-wavelengths long and the irregularities are small, they can be rearranged in sequence along the standing wave without loss of 'uniformity'. It does not matter on which of the successive circuits of the circle diagram the irregularity occurs; it only matters that it falls into its place on one of them, so that the distribution on the superimposed circuits is 'uniform'.

If the distribution is not 'uniform' in this special sense the resultant disturbance at the input is a function of the deviation from uniformity.* In practice, any type of cable will have various small irregularities in characteristic impedance along its length. In many cases the distribution of these irregularities along the cable will be random and so will approximate to a 'uniform' distribution. For any given cable, the lower the frequency the more irregularities occur within a wavelength of cable and the closer the approximation to 'uniformity' of any random distribution. For this reason, and because irregularities have only a small effect when their electrical lengths are small, the resultant disturbance at the input is very small at low radio frequencies, even for cables in which the disturbances are great at higher frequencies.

A case of special interest which illustrates the concept of 'uniformity' is that of the air-spaced coaxial cable having the inner conductor centred by a series of equispaced disc supports. Each disc support gives rise to an equal irregularity in impedance. As these irregularities are all equal and equispaced they are obviously distributed uniformly in the usual sense of the word. Consider now whether or not the distribution is uniform as defined above.

* To be more general, it is the deviation from a distribution having the special symmetry discussed in the last paragraph of Section 5 which determines the resultant disturbance. The 'uniform' distribution is the most simple one having this symmetry.

At low radio frequencies there will be many supports per half-wavelength and the irregularities will be distributed uniformly around the circle. As the frequency increases, the number in each half-wavelength decreases and the lack of 'uniformity' becomes apparent. When the supports are nearly half a wave-length apart the condition can arise where the discs in successive half-wavelengths are superimposed on the circle. At this frequency there will be a resonance, and the length of cable behaves as a stop-band filter. This variation with frequency in the properties of disc-supported coaxial lines is considered in detail in Section 4.

(4) A TRANSMISSION LINE WITH EQUISPACED IRREGULARITIES

The equispaced irregularities can take the form of a series of supporting discs, as shown in Fig. 3(a). Such a length of line
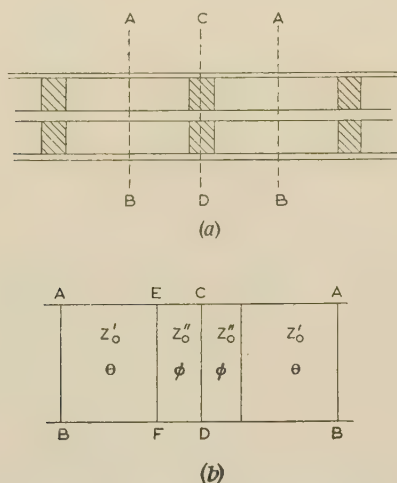
Fig. 3.—Unit quadripole for a disc-spaced line.

can be considered as a series of 4-terminal networks ABAB in cascade. Each is symmetrical and can be considered as composed of two half-sections ABCD as shown in Fig. 3(b). The part ABEF is air-spaced and has characteristic impedance Z_0' and phase length θ; in the part EFCD, which includes half the disc support, the characteristic impedance is Z_0'' and the phase length ϕ.

The effective characteristic impedance of the section ABAB will be the image impedance at AB. By consideration of the open- and short-circuit impedances it can be shown that this is given by

$$Z_0 = Z_0' \sqrt{\left(\frac{Z_0'' \tan \phi + Z_0' \tan \theta}{Z_0' - Z_0'' \tan \phi \tan \theta}\, \frac{Z_0'' \cot \phi - Z_0' \tan \theta}{Z_0' + Z_0'' \cot \phi \tan \theta}\right)} \quad . \quad (1)$$

and the phase length of the section will be

$$\beta l = 2 \arctan \sqrt{\left(\frac{Z_0'' \tan \phi + Z_0' \tan \theta}{Z_0' - Z_0'' \tan \phi \tan \theta}\, \frac{Z_0' + Z_0'' \cot \phi \tan \theta}{Z_0'' \cot \phi - Z_0' \tan \theta}\right)}$$

$$\quad \cdots \quad (2)$$

where it is assumed that the attenuation in the network is negligible.

For the A.M.R.E. Mark III coaxial line, which was used in some war-time Service equipment, the conductors were a copper tube of internal diameter $0\cdot952$ cm and a copper wire of diameter $0\cdot284$ cm. These were spaced by ceramic discs, $0\cdot50$ cm thick and $5\cdot00$ cm apart, having an effective dielectric constant of $2\cdot83$.
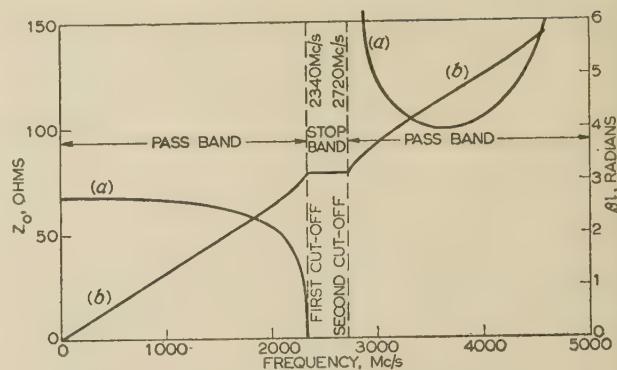
Fig. 4.—Coefficients of A.M.R.E. Mark III line at various frequencies.

(a) Characteristic impedance, Z_0.
(b) Phase length per section, βl.

In Fig. 4 are shown the calculated Z_0 and electrical length βl for a section of this line. As losses have been ignored Z_0 becomes zero at the first cut-off frequency, and with increasing frequency in the stop band becomes imaginary and increasingly large. In order to establish the order of error due to the neglect of attenuation, calculations were made with eqn. (1) appropriately modified to

$$Z_0 = Z_0' \sqrt{\left(\frac{Z_0'' \tanh \Phi + Z_0' \tanh \Theta}{Z_0' + Z_0'' \tanh \Theta \tanh \Phi}\, \frac{Z_0'' \coth \Phi + Z_0' \tanh \Theta}{Z_0' + Z_0'' \tanh \Theta \tanh \Phi}\right)}$$

$$\quad \cdots \quad (3)$$

where Θ and Φ are the complex propagation exponents in the air and ceramic sections. Assuming the resistivity of copper to be $17\cdot2 \times 10^{-6}$ ohm-cm and the power factor of the beads to be $0\cdot001$, then at a frequency as near to the cut-off as $2\,200$ Mc/s ($f/f_1 = 0\cdot94$) the effect of the attenuation would not be apparent in Fig. 4, and, at the cut-off frequencies, $Z_0 = 8\cdot72 \underline{/2°\ 28'}$, and $Z_0 = 892 \underline{/4°\ 7'}$, instead of zero and infinity. It will be evident that such corrections do not introduce any appreciable change in the curve of Z_0 against frequency.

The curve of electrical length is linear at low frequencies, but before the first cut-off frequency the electrical length increases more rapidly, remaining constant through the stop band, and then increasing rapidly at first until undisturbed values are reached.

From this example it is clear that large variations in effective transmission characteristics with frequency are possible even for a physically uniform distribution of irregularities in cases when they are not uniformly but systematically superimposed upon the impedance circle.

The above theory applies to any equispaced irregularities, and an interesting case met in practice is illustrated in Fig. 5. This Figure shows the measured variations in effective characteristic impedance of a coaxial cable which had solid polythene dielectric and extruded-lead outer conductor. Owing to a very unusual defect in the manufacturing process a series of small air bubbles were trapped between the dielectric and the lead with a regular spacing along the cable. From the measured spacing of the bubbles it was deduced that they would be $\lambda/2$ apart at a wavelength of $10\cdot2$ cm. The measured resonance was at a wavelength of $10\cdot1$ cm, which is in good agreement with the deductions made from the spacing of the bubbles.

(5) THE HARMONIC ANALYSIS OF IRREGULARITY DISTRIBUTION

A method of measuring the distribution of irregularities along a transmission line has been described by Cotte.[6,7] This method
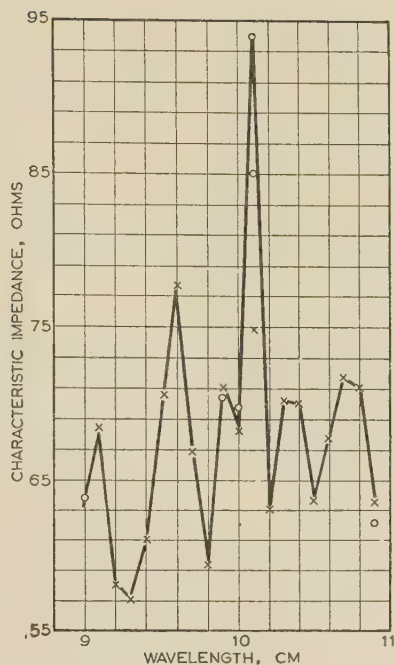
Fig. 5.—Measured variation in effective characteristic impedance for cable with equispaced air bubbles.

× Points measured by 'two-point' method.
○ Points determined from a full impedance circle.

Fig. 6.—Variation of characteristic impedance along a line.

...is discussed here in terms of the present graphical representation of the effects of irregularities partly to display the power of this method, and also because the discussion suggests variations in the test procedure which are not so evident from Cotte's analysis. In this method the variation in Z_0 along the line is considered as the sum of a Fourier series of the form $a_n \sin n\psi + b_n \cos n\psi$, where ψ is defined as $2\pi x/l$, l being the length of the cable and x the distance from the input end.

In Fig. 6(a) is shown the variation in Z_0 along a line of length l. The mean characteristic impedance of the line is Z_m, and the deviation of local Z_0 from this mean at a point x is $S(x)$. The particular variation shown in Fig. 6(a) is the sum of three harmonic terms. These are represented by the curves of Figs. 6(b), 6(c) and 6(d). In a practical case there will be many more terms, those of higher order generally being of reduced amplitude. If, as is usual, the irregularities are rounded in profile, the limit to the order of terms occurs when the wavelength of the variation in $S(x)$ becomes comparable with the extent of the individual irregularities (i.e. the 'correlation range').

The analytic device of splitting the variation into harmonic components is of no use unless these components can be measured independently. The Cotte method of doing this is to terminate the line in its mean impedance Z_m, and then the coefficients a_n and b_n of the nth term are determined by measuring the input impedance at the frequency for which the line would be $n\lambda/2$ long were it not for the irregularities. This input impedance will differ from Z_m by $\Delta_n R + j\Delta_n X$, and from this difference the coefficients can be calculated by the simple formulae

$$a_n = \Delta_n R/(n\pi) \quad b_n = \Delta X_n/(n\pi)$$

(These simple formulae only apply to lossless transmission lines. The effects of attenuation are considered by Cotte.[6,7])

Using the circle diagram, we now consider why it is that these components of the impedance difference are measures of the sine and cosine components of the nth harmonic and independent of all the other harmonics.
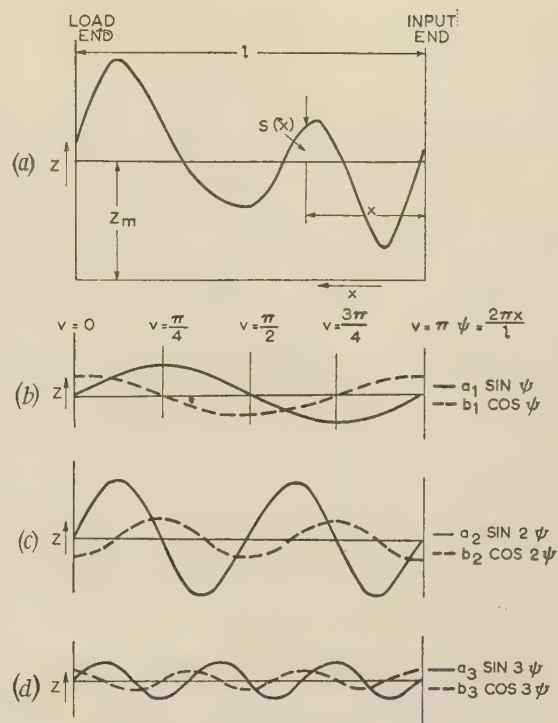
We take here a more general case than that considered by Cotte and assume that the sample of line is terminated in a pure resistance not necessarily equal to Z_m. In the following, the termination is taken to be less than Z_m, but a corresponding argument would apply if the termination were greater. Referring to the irregularity shown in Fig. 6(a), consider the measurement made at the fundamental frequency (at which the line would be $\lambda/2$ long if it were free of irregularities). The impedance changes along the line in going from the termination to the input can be displayed on one circuit of the circle diagram. The sine curve shown in Fig. 6(b) will give an increase in v.s.w.r. over the part $v = 0$ to $\pi/2$ and also an increase for the part $v = \pi/2$ to π. The effects of the two halves will thus add to give an increase in v.s.w.r. which will be apparent at the input as ΔR (a change in input resistance).

The cosine curve shown in Fig. 6(b) will have no net effect on the v.s.w.r. because the increase due to the part between 0 and $\pi/4$ will be balanced by the part between $\pi/4$ and $\pi/2$, and similarly the parts $\pi/2$ to $3\pi/4$ and $3\pi/4$ to π also cancel. Considering now the harmonic component shown in Figs. 6(c) and 6(d), it is clear that these have symmetry of distribution about 0 and π (the R-axis) and about $\pi/4$ and $3\pi/4$ (the unit circle); therefore they will make no contribution to the change of input impedance. Thus, it is clear that the change in v.s.w.r. (i.e. ΔR) is due solely to the sine component shown in Fig. 6(b), and is not influenced by any of the higher harmonics. Similarly, it can be deduced that the only component which can contribute to the phase change is the cosine component shown in Fig. 6(b).

When a measurement is made at the second-harmonic frequency the irregularities shown in Fig. 6(a) will occupy two whole circuits of the impedance circle, and it can be shown that the only components of $S(x)$ which do not have symmetry about $v = 0$ to π and $v = \pi/4$ and $3\pi/4$ are those of the second harmonic. Thus it is demonstrated graphically that Cotte's method selects at each frequency harmonic the corresponding

harmonic of $S(x)$, and that the sine and cosine components contribute to ΔR and ΔX, respectively.

The reasoning advanced above applies equally whether or not R_T, the termination of the test sample, is equal to Z_m. It has been customary when using Cotte's method to take some pains to match the cable, and fairly elaborate equalizer networks have been used in order to present a load which varies in frequency in the same way as Z_m. This accurate matching is not necessary, and the graphical approach shows that a more general theory could be developed to cover the case of any arbitrary termination, purely resistive or otherwise. A graphical solution for purely resistive terminations indicates that as the mismatch (R_T/Z_m) varies, the ratio $\Delta R/R$ remains constant. If R_T is to be chosen in order to make the best use of the available measuring equipment, it should be in the region where the percentage sensitivity in resistive measurements is greatest.

Thus it is concluded that there is no special virtue in making $R_T = Z_m$, and that R_T can be chosen to suit the measuring equipment available; also that there is no need to use anything other than a simple resistive termination even in the low-frequency range where Z_m varies considerably with frequency.

The viewpoint developed above provides a new approach to some of the points considered previously. For instance it will now be seen that the effect of an irregularity on the v.s.w.r. is greatest when the irregularity is on the unit circle ($v = \pi/4$ or $3\pi/4$) because the sin ψ component is then greatest, and the effect is zero on the R-axis ($v = 0$ or π) because there the sin ψ component is zero. Again, a short irregularity has a small effect because its components in sin ψ and cos ψ are small, and because the effects of the other terms in sin $n\psi$ and cos $n\psi$ are zero. It has been shown that, if a series of irregularities has symmetry about the R-axis and unit circle ($v = 0$ and π and $v = \pi/4$ and $3\pi/4$), it has no resultant effect. This is because there are then no terms in sin ψ and cos ψ. If the irregularities are rotated on the circle so that they now have axes of symmetry $v = v'$ and $v = v' + \pi/4$ the resultant effect will still be zero. This is because the rotation changes the phase of the components but cannot introduce the first harmonic if it is not present originally. Thus, a distribution can depart from uniformity without resultant effect if it has symmetry around the impedance circle about two v-circles differing in value by $\pi/4$.

(6) CONCLUSION

The statistical theory[8-10] of irregular transmission lines leads to the conclusion that the reflected wave at the input is dependent (*a*) upon the amplitude of the irregularities, (*b*) upon their 'correlation range', which is a measure of the length of the irregularities, and (*c*) upon the extent to which the distribution of the irregularities differs from a uniform distribution along the length of the transmission line.

The graphical approach given here leads to these same general conclusions but brings out the fact that 'uniformity' must be interpreted in a special way, namely that the distribution must be 'uniform' (or have symmetry about two v-circles which differ in phase by $\pi/4$) with respect to the standing wave on the transmission line. This concept of uniformity is not easy to introduce into a theory based on the assumption that the line carries no systematic standing wave (i.e. no standing wave other than that due to the irregularities).

On the basis of the graphical representation it is possible to explain qualitatively the increase in the reflected wave with increase in frequency in terms which are applicable, for instance, to a disc-spaced line both at low frequencies, where its discs can be considered to be uniformly spaced, and at higher frequencies where the disc structure resonates.

A qualitative explanation of Cotte's method of measuring the distribution of irregularities along a line can be given. From this it is concluded that there is no necessity for the accurate matching of the test sample by the use of equalizer networks and that the termination can be chosen to suit the characteristics of the measuring equipment.

(8) REFERENCES

(1) JACKSON, WILLIS: 'Radio-Frequency Transmission Lines' (Methuen, 1945).
(2) JACKSON, WILLIS, and HUXLEY, L. G. H.: 'The Solution of Transmission-Line Problems by use of the Circle Diagram of Impedance', *Journal I.E.E.*, 1944, **91**, Part III, p. 105.
(3) SLATER, J. C.: 'Microwave Transmission' (McGraw-Hill, 1942).
(4) SMITH, P. H.: 'Transmission-Line Calculator', *Electronics*, January, 1939, **12**, p. 29.
(5) DE ONIS, A.: 'A Graphic Procedure for Calculations Involving Transmitting Line Systems', *Communication News*, 1949, **10**, p. 87.
(6) COTTE, M.: 'La détermination des écarts d'impédance locaux dans un câble et de la fonction de corrélation de ces écarts', *Bulletin de la Société Française des Électriciens*, 1948, **8** (Sér. 6), p. 16.
(7) COTTE, M.: 'Étude d'une longueur de fabrication de câble par des mesures d'impédance terminale', *Câbles et Transmission*, 1955, **9**, No. 2, p. 161.
(8) DIDLAUKIS, M., and KADEN, H.: 'Die inneren Ungleich-mässigkeiten von koaxialen Breitbandkabeln', *Elektrische Nachrichten-technik*, 1937, **14,** p. 13.
(9) BRILLOUIN, L. N.: 'Irregularities in Telephone and Television Coaxial Cables', *Electrical Communication*, 1938, **17**, p. 164.
(10) ROSEN, A.: 'The "Unit" Treatment of Impedance Irregularities and its Application to Long Lines', *Proceedings I.E.E.*, Monograph No. 98, May, 1954 (**101**, Part IV, p. 271).

ELECTRIC FIELD STRENGTH BETWEEN CROSSED AND PARALLEL CIRCULAR CYLINDERS

By P. G. HARPER, B.Sc., Ph.D., and J. J. O'DWYER, B.Sc., B.E., Ph.D., A.Inst.P.

SUMMARY

An approximate method is given for evaluating the maximum field strength between charged circular cylinders whose axes are either parallel or perpendicular. The approximate formula is compared with exact results and shows an error of less than $\frac{1}{2}\%$ for a separation of one diameter. If the error in the approximate formula for crossed cylinders is of the same order, then the difference between the results for the crossed and parallel cases is significantly greater than this error.

LIST OF SYMBOLS

a = Nearest inter-electrode distance, cm.
E_{max} = Maximum electric field strength in the gap, volts/cm.
R = Radius of the cylindrical electrode, cm.
V_0 = Potential difference between electrodes, volts.
$V(x, y, z)$ = Potential of the field at any point whose co-ordinates are x, y, z, volts.
η = Utilization factor.

(1) INTRODUCTION

In high-voltage work, particularly in connection with the development of spark-gaps, the problem naturally arises of finding the maximum field strength, E_{max}, associated with a given electrode configuration, the potential difference being V_0. This is expressed empirically by

$$E_{max} = \frac{V_0}{a} \frac{1}{\eta} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where a is the shortest inter-electrode distance, and η, which depends only on the geometry, is called the utilization factor. Exact calculations have been made for η with sphere-gaps[1] and parallel circular cylinders,[2] and it has been found experimentally that crossed circular cylinders give the same results as parallel cylinders.[3] It is the purpose of the paper to derive power-series expansions for the utilization factor of parallel equal[2] circular cylinders and circular cylinders at right angles. It is found that the first three terms of the series give a result whose accuracy is sufficient for all practical purposes for separations up to the cylinder diameter. It is also found that crossed and parallel cylinders differ by a term which is small.

(2) EXACT SOLUTION FOR PARALLEL CYLINDERS

Consider two parallel lines, distant $2b$ apart, carrying equal line charge densities of opposite sign. Then, as is well known, the potential function is given by

$$V = \mathscr{R}\left\{ C \log \left[(\zeta - b)/(\zeta + b) \right] \right\} \quad . \quad . \quad (2)$$

where ζ is the complex variable, and C is a constant determined by the line charge density. The equipotentials given by eqn. (2) are families of parallel circular cylinders, and the field distri-

bution for two such charged cylinders is found by replacing the appropriate two equipotentials by conducting surfaces. It is easily shown that for two cylinders each of radius R, the maximum field strength is given by

$$E_{max} = \frac{V_0}{a} \frac{\sinh \left[V_0/(2C) \right]}{V_0/(2C)} \quad . \quad . \quad . \quad (3)$$

where C is found from

$$a/R = 4 \sinh^2 \left[V_0/(4C) \right] \quad . \quad . \quad . \quad (4)$$

This result holds for all values of a/R and is simpler than the usual expression of the same quantity.[2]

(3) APPROXIMATE SOLUTION FOR PARALLEL CYLINDERS

The method to be used consists in expanding E_{max} in a power series in a/R. Accordingly we take a Cartesian system of axes and consider two circular cylinders whose axes are parallel to the z-axis and intersect the y-axis at a distance $R + a/2$ above and below the origin. The equations of the cylinders are

$$\left[y \pm \left(\frac{a}{2} + R \right) \right]^2 + x^2 = R^2 \quad . \quad . \quad . \quad (5)$$

and these can be expanded by the binomial theorem so that adjacent surfaces are represented by

$$y = \pm \left(\frac{a}{2} + \frac{x^2}{2R} + \frac{x^4}{8R^3} \cdots \right) \quad . \quad . \quad . \quad (6)$$

these three terms being required to carry the approximation to the second power in a/R.

If we now represent the potential by a polynomial $V(x, y, z)$ it is found that a fifth-order expression will give a consistent approximation. The general fifth-order polynomial can be greatly simplified by observing that

    (a) V is independent of z
    (b) V is an even function of x
    (c) V is an odd function of y

after which the only remaining terms are

$$V = Ay^5 + y^3(Bx^2 + C) + y(Dx^4 + Fx^2 + G) \quad . \quad (7)$$

Application of Laplace's equation to ensure that (7) really is a solution of the field equations eliminates three of the constants so that

$$V = Ay^5 - y^3(10Ax^2 - C) + y(5Ax^4 + 3Cx^2 + G) \quad (8)$$

It is further required that

$$V \equiv \pm V_0/2$$

when on the conducting surfaces given by the expansion (6). Because of symmetry the relations obtained for the two surfaces are the same, so we need consider only one of them. Substituting (6) in (8) and equating coefficients of powers of x up to and including the fourth, we find three equations for the three

unknowns A, C and G for which solution to the required power yields

$$A = \frac{V_0}{a} \frac{1}{5a^2R^2}$$

$$C = \frac{V_0}{a} \frac{1}{3aR}\left(1 - \frac{a}{3R}\right) \qquad \cdots \quad (9)$$

$$G = \frac{V_0}{a}\left(1 - \frac{a}{12R} + \frac{11a^2}{720R^2}\right)$$

From (7), however, the maximum field strength is given by

$$E_{max} = -\left(\frac{\partial V}{\partial y}\right)_{\substack{x=0 \\ y=a/2}}$$

$$= -\left(\frac{5}{16}Aa^4 + \frac{3}{4}Ca^2 + G\right)$$

$$= -\frac{V_0}{a}\left(1 + \frac{a}{6R} - \frac{a^2}{180R^2}\cdots\right) \quad \cdots \quad (10)$$

by virtue of (9). This is an interesting power series since its coefficients give very rapid convergence even for values of a/R which are not small. Since (3) and (4) give the exact solution to this problem we can check the accuracy of (10) for any value of a/R.

Table 1

ACCURACY OF SERIES EXPANSION (10)

a/R	1	2	3	4
Error in expansion (10)	% 0·052	% 0·32	% 0·83	% 1·7

Some results are shown in Table 1, the series in all cases giving too low a value for E_{max}. Higher values of a/R are of less interest since corona occurs for $a/R > 3·85^2$.

(4) APPROXIMATE SOLUTION FOR CROSSED CYLINDERS

Consider in this case two circular cylinders, one parallel to the x-axis and the other parallel to the y-axis, with the axes of the cylinders intersecting the z-axis at a distance $R + a/2$ above and below the origin. Then, as above, the expansions for the adjacent surfaces are

$$z = \frac{a}{2} + \frac{x^2}{2R} + \frac{x^4}{8R^3}$$
$$z = -\left(\frac{a}{2} + \frac{y^2}{2R} + \frac{y^4}{8R^3}\right) \qquad \cdots \cdots \quad (11)$$

The conditions to be laid on the polynomial potential function $V(x, y, z)$ are then

 (a) V is an even function of both x and y.
 (b) $V(x, y, z) = -V(y, x, -z)$.

The second condition arises from the fact that rotation of each cylinder through a right angle and inversion of the z-axis reproduce the original situation with the sign of the potential reversed. These conditions, together with the application of Laplace's equation, lead to a potential

$$V = Az^5 - (5Ax^2 + 5Ay^2 - C)z^3$$

$$+ \left[Dx^4 + Dy^4 - (6D - 15A)x^2y^2 - \frac{3C}{2}(x^2 + y^2) + F\right]z$$

$$- 6M(x^2 - y^2)z^2 + Mx^4 - My^4 + Nx^2 - Ny^2 \quad (12)$$

From this the maximum field strength is given by

$$E_{max} = -\left(\frac{\partial V}{\partial z}\right)_{\substack{x=y=0 \\ z=a/2}}$$

$$= -\left(\frac{5}{16}Aa^4 + \frac{3}{4}Ca^2 + F\right) . \quad \cdots \quad (13)$$

Substitution of (11) in (12) to determine the boundary conditions gives, on equating the coefficients of powers and products of x and y up to the fourth, six equations for the six unknowns A, C, D, F, M and N. Upon solution we find, to the required order,

$$A = \frac{V_0}{a} \frac{1}{10a^2R^2}$$

$$C = \frac{V_0}{a} \frac{1}{3aR}\left(1 - \frac{a}{12R}\right) \qquad \cdots \quad (14)$$

$$F = \frac{V_0}{a}\left(1 - \frac{a}{12R} + \frac{a^2}{1\,440R^2}\right)$$

Substitution of (14) in (13) gives

$$E_{max} = -\frac{V_0}{a}\left(1 + \frac{a}{6R} - \frac{a^2}{90R^2}\cdots\right) \quad \cdots \quad (15)$$

Thus the maximum field strengths for crossed and parallel cylinders differ only in the second-order term, so that experimental results may be indistinguishable for small relative separation. In fact, the difference between the results for crossed and parallel cylinders as given by the approximations (10) and (15) is shown in Table 2. It must, however, be stressed that

Table 2

DIFFERENCE BETWEEN SERIES EXPANSIONS (10) AND (15)

a/R	1	2	3	4
Difference between crossed and parallel approximations	% 0·48	% 1·7	% 3·4	% 5·5

there is no way of checking the absolute accuracy of (15) and it can only be guessed that the error is of the same order of magnitude as that of (10).

Finally it should be mentioned that from the form of the results it could be expected that

$$E_{max} = -\frac{V_0}{a}\left[1 + \frac{a}{6R} - (1 + \sin^2\theta)\frac{a^2}{180R^2}\cdots\right]$$

for cylinders whose axes are inclined at an angle θ.

(5) ACKNOWLEDGMENT

(6) REFERENCES

(1) RUSSELL, A.: 'A Treatise on the Theory of Alternating Currents' (Cambridge University Press, 1904).
(2) PEEK, F. W.: 'Dielectric Phenomena in High-Voltage Engineering' (McGraw-Hill Book Co., 1929).
(3) SCHWAIGER, A., and SORENSEN, R. W.: 'Theory of Dielectrics' (John Wiley and Sons, 1932).

FUNDAMENTAL RADIO SCATTER PROPAGATION THEORY

By the late E. C. S. MEGAW, M.B.E., D.Sc., Member.

PREFACE

The paper presents an account of the theoretical investigations of radio scatter propagation in the region well beyond the horizon, and of the related phenomena of stellar scintillation, which were carried out by Dr. Megaw prior to his sudden death. He had been considering the publication of such a work, partly to amplify his own earlier published (though abbreviated) accounts, but mainly to provide a fundamental interpretation of the physical problems involved. The present paper has been compiled from his extensive notes and analytical material in an attempt to fulfil this purpose. Any limitations which may be apparent in the presentation are the responsibility of his two colleagues who prepared the present text.

F. A. K.
M. A. J.

SUMMARY

The paper presents a detailed theoretical treatment of the beyond-horizon radio scatter propagation problem. The form for the spectral density of refractive-index fluctuation, assumed to be the same as that of velocity fluctuation, is derived from the universal equilibrium theory of turbulence. It is used to predict the spectral densities and magnitudes of brightness fluctuation and angular deflection fluctuation for light from a stellar source, and of field-strength fluctuation for line-of-sight radio propagation. These predictions are compared with experimental data, and support is thereby obtained for the theoretical form of the refractive-index fluctuation spectrum. The height variation of the intensity of the refractive-index fluctuations is derived from the experimental data. There follows a detailed treatment of the beyond-horizon case of radio scatter propagation, including a full analysis of the influence of the geometry of the scattering volume on the received field-strength. It is shown that the form of the scattering cross-section implied in the results is compatible with those obtained in different treatments, based on idealized assumptions.

LIST OF PRINCIPAL SYMBOLS

a = Eddy size, defined as $\pi/$(wave number k).

$a_1 = \pi/k_1$.

a_1' = Maximum eddy size used for small-angle scattering formula for metre wavelengths. (Section 4.5.3.)

a_2 = Maximum eddy size assumed to contribute to field strength for beyond-horizon propagation.

a_m = Adopted transitional eddy size between diffraction and ray-theory treatments for line-of-sight radio propagation.

a_{min} = Minimum eddy size for starlight-bandwidth discussion.

$a_s = 5 \cdot 9 \, (\nu^3/\varepsilon)^{1/4}$.

A = Telescope aperture.

$A(k)$ = Complex Fourier transform of Δn.

b = Ratio ϕ_1'/θ_1' of mean azimuth and elevation beam angles.

B = Numerical constant occurring in the expression (75) for the scattering cross-section in the beyond-horizon case.

c = Constant occurring in Reference 15.

C = Quantity occurring in the integrand of the volume integral for beyond-horizon radio scattering
$= \theta \tan(\psi - \alpha/2 - \theta)$

d_0 = Sum of the radio horizon distances d_1 and d_2 of transmitter and receiver
$= \sqrt{(2Rh_1)} + \sqrt{(2Rh_2)}$

D = Function of α, θ_1', ϕ_1' occurring in the result of the volume integration for beyond-horizon case.

$\tilde{E}$ = R.M.S. electric-field fluctuation, integrated over volume (diffraction) or path (ray theory) and spectrum.

E_0 = Free-space field strength at a point at a distance from the source equal to that of the point of observation.

E_i = Free-space field strength due to the source incident on the Fourier element under consideration.

$\tilde{E}_k$ = R.M.S. electric-field fluctuation, integrated over volume or path, for wave number k.

E_s = Scattered field produced by passage through a Fourier element ($= E_{s0}$ for $\beta = \pi/2$, $\gamma = 0$).

$E_\epsilon(k)$ = Spectral density of mean-square permittivity fluctuation in Reference 6.

f = Frequency corresponding to an eddy, taken as being equal to the reciprocal of its period.

$F(k)$ = Fourier transform in vector k in wave-number space of mean-square permittivity fluctuation in Reference 6.

$\mathscr{F}_n(k)$ = Spectral density in vector k of mean-square fluctuation.

$F_n(k)$ = Three-dimensional relative spectral density in terms of wave-number magnitude of mean-square refractive-index fluctuation.

$F_{nx}(k_x)$ = One-dimensional spectrum derived from $F_n(k)$.

g_{11}, g_1 = Scalar correlation functions for longitudinal and transverse fluctuating velocity components.

G = Assumed ratio of mean refractive-index gradient at height h to that at sea level.

h_1, h_2 = Aerial heights (subscripts 1 and 2 refer to transmitter and receiver respectively).

H = Maximum height at which refractive index fluctuation intensity is appreciable.

J_y = Displacement current density in the y-direction at a point in a Fourier element.

J = Total dielectric current over the central cross-section perpendicular to the electric vector of a Fourier element.

k = Wave number of an eddy [$\pi/$(eddy wavelength)].

k_1 = Wave number at which simple form of spectral density must be cut off to give same total energy as the actual spectrum.

$k_1', k_2, k_m, k_{min}, k_{\gamma_1}, k_{\gamma_2} = \pi/a_1', \quad \pi/a_2, \quad \pi/a_m, \quad \pi/a_{min}, \quad \pi/a_{1\gamma}, \quad \pi/a_{\gamma_2}$, respectively.

$k_s = (3c^2/8)^{1/4}(\epsilon/\nu^3)^{1/4}$ = wave number around which effect of viscosity becomes important.

k_x = Wave number in the 1-dimensional spectrum.

$K = |k_0 - k_1| =$ wave number for which the Fourier transform is required for simple formula for scattering cross-section in Reference 6.

$l =$ Integral scale $=$ infinite integral of the correlation function of a scalar field quantity.

$l_{11}, l_1 =$ Respective integral scales for the correlation functions g_{11}, g_1 for a vector field.

$l_s =$ Cut-off scale introduced in exponential correlation function in Reference 12.

$L =$ Path length from transmitter to receiver.

$L_k =$ Distance $(2a^2/\lambda)$ of transition for the eddy size a from ray-theory to diffraction treatment.

$m =$ Ratio $(\Delta\lambda/\lambda_0)$ in the discussion of the effect of light bandwidth.

$n =$ Refractive index of the atmosphere.

$\bar{n}', \bar{n}_L =$ Height gradient of mean refractive index for radio and light waves respectively.

$N =$ Fluctuating component of Δn (r.m.s. value $\widetilde{N}$).

$\bar{N} =$ Mean component of Δn.

$\widetilde{N}', \widetilde{N}_L =$ R.M.S. values over the spectrum of (N_k/a) for radio and light waves respectively.

$N_k = [F_n(k)dk]^{1/2}$.

$(N_k)_s =$ Value of N_k at sea level [suffix omitted in Section 4.3 after eqn. (43)].

O: symbol meaning 'of the same order as', e.g. $D_1 = O(D2)$.

$p = 2\theta/\alpha$.

$P =$ Local atmospheric pressure in millibars.

$Q =$ Geometrical factor in eqn. (58).

$r =$ Scalar distance.

$r_1, r_2 =$ Distances of the Fourier element under consideration from the source and the point of observation respectively.

$\delta r =$ Increment of length along ray path.

$R =$ Effective radius for radio propagation of the earth.

$s =$ Local partial pressure of water vapour, in millibars.

$S =$ Geometrical factor in eqn. (65).

$T =$ Absolute temperature.

$t =$ Period associated with an eddy.

$t_1, t_s =$ Eddy periods corresponding to the eddy sizes a_1, a_s, in the scintillation discussion.

$\tilde{u} =$ R.M.S. fluctuating velocity component.

$U =$ Assumed mean horizontal wind velocity in the troposphere.

$v =$ Volume, in particular, scattering volume.

$\alpha =$ Angle between the tangent planes to the earth (assumed spherical) from transmitter and receiver $= LR^{-1}(1 - d_0/L)$.

$\beta =$ Angle between incident electric field and direction of scattering.

$\gamma =$ Angle between direction of incidence and direction of scattering.

$\gamma_a =$ Half-energy beam width $(\sim\lambda/2a)$ of scattered main lobe from a Fourier element for $L \gg L_k$, $a \gg \lambda$.

$\gamma_{max} =$ Maximum value of γ over the scattering volume for the metre-wave case.

$\Gamma =$ Total angular deviation of light ray.

$\Gamma_y, \Gamma_z =$ Components of Γ.

$\Delta =$ Linear deviation of light ray.

$\Delta_y, \Delta_z =$ Components of Δ.

$\Delta n = n - 1$.

$\Delta\lambda =$ Bandwidth of starlight.

$\varepsilon =$ Local mean rate per unit mass of kinetic-energy dissipation in the turbulence. Also used for the base of natural logarithms.

$\delta\epsilon =$ Permittivity deviation (Reference 6).

$\epsilon_0 =$ Permittivity of free space.

$\zeta =$ Zenith distance.

$\eta =$ Dissipation length $[= (\nu^3/\varepsilon)^{1/4}]$.

$\theta =$ Elevation angle from the transmitter.

$\theta_1, \theta_2 =$ Elevation semi-beam widths of the transmitting and receiving aerials.

$\theta'_1 = (\theta_1\theta_2)^{1/2}$.

$\lambda =$ Radiation wavelength.

$\lambda_0 =$ Critical bandwidth of starlight, taken as a_s^2/L, above which an appreciable effect is produced.

$\mu_1^2, \mu_2, \mu_3^2 =$ Multiplying factors for the simple spectral density respectively produced by the extension to wave numbers outside the inertial sub-range, and the effects of instrumental time-constant and the aperture of the telescope.

$\nu =$ Kinematic viscosity.

$\xi = h/H$.

$\rho =$ Correlation function of a scalar field quantity; also air density.

$\sigma =$ Scattering cross-section.

$\tau =$ Instrumental time-constant.

$\phi =$ Azimuth angle from transmitter, measured from direction transmitter-receiver.

$\phi_1, \phi_2 =$ Azimuth semi-beam widths of the transmitting and receiving aerials.

$\phi'_1 =$ The lesser of ϕ_1, ϕ_2.

$\chi =$ Angle between transmitter-receiver and direction of scattering for line-of-sight diffraction case.

$\psi =$ Angle in vertical plane containing transmitter and receiver, measured from the vertical at the common intersection of this plane and the horizon planes through transmitter and receiver.

$\omega =$ Radiation frequency.

(1) INTRODUCTION

The rapid progress in knowledge of beyond-horizon radio scatter propagation during the past few years has provided a good example of the truth of the well-known remark that 'there is nothing so practical as a good theory'. Indeed, even an imperfect theory often proves much more practical than none at all; such a theory enabled predictions of the possibilities and of the broad characteristics of tropospheric scatter propagation of radio waves to be made by the author in early 1949 (in a restricted publication), and these were subsequently confirmed experimentally.[1] The basis of the theory was that the lower atmosphere is permanently turbulent, and that the resulting fluctuations in refractive index provide the explanation of persistent radio propagation to distances well beyond the radio horizon. In the absence of adequate meteorological knowledge, the characteristics of the refractive-index fluctuations were predicted from the universal equilibrium theory of turbulence, supported by analysis of the limited stellar scintillation data available. These results are presented in detail in the present paper, and applied in a complete treatment of the radio scatter problem.

An independent theory of the long-distance radio scatter propagation phenomenon was presented by Booker and Gordon[2] in 1950, adapting for the purpose a theory of scattering for scalar fields given by Pekeris.[3] Booker and Gordon used a different representation of the refractive-index fluctuations in the atmosphere, with a spectrum corresponding to an exponential correlation function, and a more rigorous basic approach than in the present treatment; however, the error thus caused in the latter can be estimated and allowed for, and advantage then taken of the attempt in the present treatment to take account of all effects likely to be important in practice.

More recent treatments of the theory of radio scattering by turbulence have been given by Villars and Weisskopf.[4, 5] The same spectrum of atmospheric turbulence is used as in the present paper, but different methods of deduction of the statistics of the refractive-index fluctuations are employed. On the hypothesis used by Villars and Weisskopf in their first paper,[4] i.e. that the refractive-index fluctuations were caused by the pressure fluctuations brought about by the turbulence, it is necessary to postulate values of the r.m.s. turbulent velocity which are considerably too large. In fact, as has been shown by Batchelor,[6] and Gallet,[7] the refractive-index fluctuations are likely to be principally caused by turbulent convection in the presence of a spatial gradient of mean refractive index. However, Villars and Weisskopf's revised hypothesis[5] on these lines appears to be invalid, on account of confusion regarding the definition of an 'eddy'; this is discussed presently. When their final result is corrected, the wavelength dependence disappears. This reduces the agreement with experiment reported by Norton et al.[8]

(2) GENERAL DISCUSSION OF EDDIES, SPECTRA, SCALES AND CORRELATIONS

Before proceeding to the discussion of the particular statistical properties of the turbulence and the resulting refractive-index fluctuations, and hence the scattering of light and radio waves to be expected, some brief general remarks on the statistical representation of fluctuating field quantities are required. For a fuller discussion reference may be made to Batchelor's work.[9]

In problems involving fluctuating field quantities, much is to be gained by using the spectral density explicitly in the analysis (cf. Batchelor[9]). The spectral density in scalar wave-number k of mean-square refractive-index fluctuation will be written as

$$\widetilde{\Delta n}^2 F_n(k) \quad . \quad . \quad . \quad . \quad (1)$$

where $\widetilde{\Delta n}$ is the total r.m.s. fluctuation of Δn ($= n - 1$), which is assumed to be a stationary random function of space and time, and the 'relative power spectrum' F_n, in terms of wave-number magnitude, represents an average taken over a sufficiently long time.

If $A(k)$ is the complex Fourier component of Δn, defined over a large volume v, for the vector wave-number k, the spectral density in wave-number space of the mean-square fluctuation is given[6] by

$$\mathscr{F}_n(k) = \frac{\overline{A(k)A^*(k)}}{8\pi^3 v}$$

the asterisk denoting the complex conjugate. (It is this spectral density which, by the Wiener–Khintchine theorem, is the Fourier transform of the auto-covariance of Δn in physical space.)

The assumption that Δn is isotropic, so that $\mathscr{F}_n(k)$ is independent of the direction of k, is now introduced. It is then possible to integrate over all directions of k and obtain the relative spectral density in terms of wave-number magnitude, i.e.

$$F_n(k) = \frac{4\pi k^2 \mathscr{F}_n(k)}{\widetilde{\Delta n}^2}$$

It may be noted that it is related to the functions $E_\epsilon(k)$ and $F(k)$ used by Batchelor,[6] thus:

$$E_\epsilon(k) = 4\pi k^2 F(k) = 4\widetilde{\Delta n}^2 F_n(k) \quad . \quad . \quad . \quad (2)$$

since $\widetilde{\delta\epsilon}^2 = 4\widetilde{\Delta n}^2$; normalization constants of order unity may be multiplied into $F_n(k)$.

F_n necessarily satisfies the relation

$$\int_0^\infty F_n(k)dk = 1 \quad . \quad . \quad . \quad . \quad (3)$$

If the value of Δn is measured along a straight line in space and the result is Fourier-analysed, numerically or electrically, the result obtained (subject to proper precautions), is a one-dimensional spectrum $F_{nx}(k_x)$, in which, for a given k_x, contributions come from all $k > k_x$ in the 3-dimensional spectrum. For a scalar field quantity, such as Δn (cf. Kovasznay et al.[10]),

$$F_{nx}(k_x) = \int_{k_x}^\infty k^{-1}F_n(k)dk \quad . \quad . \quad . \quad (4)$$

For a transverse vector field quantity such as velocity (in an incompressible fluid) there is an additional factor $(1 - k_x^2/k^2)$ in the integrand.

(2.2) Correlation Functions and Scales

The correlation function $\rho(r)$, i.e. the normalized covariance, of a scalar field quantity and the one-dimensional relative spectrum function $F_{nx}(k_x)$ are each other's Fourier cosine transforms. The same relationship holds for the turbulent velocity components, but there are two distinct (scalar) correlation functions, $g_{11}(r)$ and $g_1(r)$, for the longitudinal and transverse components, respectively (cf. Batchelor[9]). If the Fourier pair is written so as to satisfy the normalization condition corresponding to eqn. (3), as in Batchelor[9] and Megaw[11] (but not in Kovasznay et al.[10] and a number of other papers),

$$F_{nx}(0) = \frac{2l}{\pi} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where l is the integral scale defined by

$$l = \int_0^\infty \rho(r)dr \quad . \quad . \quad . \quad . \quad (6)$$

and similarly for the two integral scales of the velocity field ($l_{11} = l_1$ subject to isotropy).

These three scales l are all of the same order of magnitude, corresponding to that of the size of the larger eddies into which energy is fed from some external source in an approximately stationary field of turbulence. This order of magnitude can be estimated from measured correlations, by fitting roughly to an assumed exponential form, or otherwise. However, little more can be learnt from such measurements; the apparently small discrepancies between, say, a simple exponential and the true curve measured as a long-period average are of vital importance in the region where the correlation is near unity.[6]

A pair of values for $\widetilde{\Delta n}$ and l, with some convenient assumption about the shape of the correlation, is insufficient for the determination of the scattered field. Nor are two recently-suggested modifications[12, 13] of the exponential correlation function, which respectively include a cut-off at a scale $l_s \ll l$ and an anisotropic correlation, altogether satisfactory. It appears preferable to work directly in terms of the spectrum function, introducing a smallest eddy size which is a direct measure of the point at which the spectrum is cut off by viscosity.

(2.3) Definition of an Eddy

The notion of an 'eddy' can be defined in two ways, namely

(a) An entity associated with a length of order $(1/k)$ (taken as π/k in the present paper) corresponding to a wave-number magnitude k in a 3-dimensional Fourier analysis, and a (relative) fluctuation energy $F_n(k)dk$.

(b) Exactly as in (a), but associated with a (relative) fluctuation energy

$$\int_k^\infty F_n(k)dk. \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

In other words, echoing L. F. Richardson's rhyme,* an eddy of type (a) is a 'whirl' of size π/k; an eddy of type (b) is the same 'whirl' carrying with it all its 'lesser whirls, and so on to viscosity'. Villars and Weisskopf made their hypothesis of turbulent mixing [Reference 5, eqn. (28)] in terms of an eddy of type (b). However, the 'little whirls', whose integrated effect is allowed for in the mathematics, are by hypothesis uncorrelated, and so make no net contribution to convection over the length π/k. It appears that a hypothesis appropriate to eddy (a) has been injected into an analysis in terms of eddy (b); the result is that the index of L in the final answer, eqn. (36), is one unit too high. On making this change, the spectral density proportional to k^{-2}, corresponding to the exponential correlation, and therefore the scattering formula given by Booker and Gordon is recovered. The wavelength dependence of eqn. (44), which Norton et al.[8] considered to fit the experimental results, then disappears, and, instead of eqn. (46), the actual scattering cross-section implied by Megaw's result[1] is obtained.

(3) REFRACTIVE INDEX FLUCTUATIONS IN THE ATMOSPHERE

(3.1) General

It is well known that the refractive index, n, of the lower atmosphere at radio wavelengths may be represented approximately by the expression

$$n - 1 = \frac{79 \times 10^{-6}}{T}\left(P + 4800\frac{s}{T}\right) \quad . \quad . \quad . \quad (8)$$

At light wavelengths the water-vapour contribution to the refractive index is about 20 times smaller.

Consideration will be restricted to conditions in which mean refraction may be represented by an effective earth radius of the same order as the so-called standard value (4/3 of the real radius), which corresponds approximately to the long-term mean distribution. Since radio scattering is practically important precisely when there is little superrefraction, this limitation is not serious. The quantity Δn is therefore represented by the sum of a steady component and a superimposed fluctuation, i.e.

$$\Delta n \equiv n - 1 = \bar{N} + N \quad . \quad . \quad . \quad (9)$$

The component $\bar{N}$, which is of order 10^{-4}, is thus a slowly decreasing function of height alone, and is steady in the sense that its time variation is negligible over periods long in comparison with those of the fluctuations.

Available meteorological knowledge, while providing useful guidance as to orders of magnitude, is not sufficiently detailed to provide statistical properties of N, even at low levels. Information about the refractive-index fluctuations was therefore sought from a comparison of observational data on the scintillation of light waves from stars with the results of a theoretical treatment, which is similar to the treatment of radio scattering to be given. Hence the presentation of both treatments, and of the deductions about the refractive-index fluctuations, is deferred until Section 4.

(3.2) Spectrum within the Inertial Sub-Range

A theoretical prediction of the form of the refractive index spectrum is now described. The important step is to relate the

* 'Big whirls have little whirls
    That feed on their velocity;
 And little whirls have lesser whirls—
    And so on to viscosity.'

spectral density of mean-square refractive-index fluctuation with that of mean square turbulent-velocity fluctuation, the theoretical prediction for which is well known and reasonably consistent with experimental results. The most plausible assumption is that the two spectra are of the same form. Batchelor has recently discussed this problem and arrived at the same conclusion. He showed that pressure changes in both the troposphere and the ionosphere, resulting directly from the turbulent motion, were probably unimportant in producing changes of n (except for pressure changes on so large a scale, and hence occurring so slowly, that their effect would not be called scintillation). This conclusion is also reached by Villars and Weisskopf in their latest paper,[5] on account of the failure of their original theory, based on pressure-change considerations, to explain the observed levels of scattered power. Thus the refractive-index fluctuations may be regarded as being directly caused by fluctuations of temperature (or water-vapour concentration) in the lower atmosphere, and of electron density in the ionosphere, which are themselves produced by turbulent mixing.

Batchelor concluded that the contribution to the spectral density $E_\epsilon(k)$ (his notation) from each quantity would be of similar form, and that

$$E_\epsilon(k) \simeq \widetilde{\Delta}n^2 l^{-2/3}k^{-5/3} \quad . \quad . \quad . \quad (10)$$

The corresponding normalized spectral density (notation of the present paper) is shown by substitution in eqn. (3) to be

$$F_n(k) = \frac{2}{3}k_1^{2/3}k^{-5/3} \quad . \quad . \quad . \quad . \quad (11)$$

Eqns. (10) and (11) are valid for all values of k within the 'inertial' sub-range, i.e. satisfying $(\varepsilon/\nu^3)^{1/4} \gg k \gg k_1$. They have the same form as the expression for the spectral density of mean square velocity fluctuation in the inertial sub-range derived from the universal equilibrium theory of turbulence. The wave number k_1 is defined as the wave number at which a spectrum with the spectral density assumed to be proportional to k^{-5} for all $k \ll k_s$ (the actual value of k_s has negligible effect in this connection)—would have to be cut off to give the same total energy as the actual spectrum.[14]

(3.3) Spectrum outside the Inertial Sub-Range

An expression for $F_n(k)$ for wave numbers outside the inertial sub-range will sometimes be required. Defining the symbol O to indicate 'of the same order of magnitude as', approximate results for the cases $k = O(k_1)$ and $k = O(k_s)$ are now derived [In applications, little error is involved in assuming that these results apply for all $k \leqslant O(k_1)$, and $k \geqslant O(k_s)$, respectively]. For the large-eddy end of the spectrum, if the whole field of turbulence is assumed to be homogeneous and isotropic, physical reasoning suggests that, as the quantity $K \equiv (4\pi/\lambda) \sin \gamma/2 \to 0$, the scattering cross-section, σ, will go over to the 'Rayleigh scattering' form, i.e.

$$\sigma = (\text{constant})\, \widetilde{\Delta}n^2 l^3 \lambda^{-4} \quad . \quad . \quad . \quad (12)$$

It will be shown later that $\sigma \propto F_n(K)$, whence

$$F_n(k) = (\text{constant})\, k^{-1}(k/k_1)^2 \quad . \quad . \quad . \quad (13)$$

The following expression will be used for all $k \ll k_s$ [the normalization factor 2/3 of eqn. (11) requires revision in view of this extension of the range of k; numerical integration of (14) and substitution in eqn. (3), gives 0·80 as the appropriate figure]

$$F_n(k) = 0\cdot80 k_1^{-1}\left(\frac{k}{k_1}\right)^2\left[1 + \left(\frac{k}{k_1}\right)^2\right]^{-11/6} \quad . \quad (14)$$

this interpolation formula for $F_n(k)$ is consistent with both eqns. (11) and (13). In practice, the large-eddy end of the turbulence spectrum is not isotropic, so that eqn. (14) is only approximate. It is not, of course, implied that eddies of sizes $a \gg a_1$ actually exist;* eqn. (14) is, however, required in order to estimate σ when γ is very small and

$$\frac{4\pi}{\lambda}\sin\frac{\gamma}{2} = K \ll k_1$$

For the exponential correlation $\rho(r) = \varepsilon^{-r/l}$,

$$F_n(k) = \frac{2l}{\pi}\frac{2k^2l^2}{(1+k^2l^2)^2} \quad \cdots \quad (15)$$

This result is obtained from the 1-dimensional relative spectral density $F_{n_x}(k_x)$ by inverting eqn. (4) to give

$$F_n(k) = (\text{constant})\left[-k\frac{\partial F_{n_x}(k_x)}{\partial k_x}\right]_{k_x=k} \quad \cdots \quad (16)$$

By obtaining the 1-dimensional spectrum $F_{n_x}(k_x)$ corresponding to eqn. (14), the integral scale of the refractive-index fluctuations can be shown to be equal to $0\cdot24a_1$, where $a_1 = \pi/k_1$. The integral scale of the exponential correlation is l, which may therefore be put approximately equal to $1/k_1$. The dependence of eqn. (15) on k then differs from that of eqn. (14) only in the slight change from 11/6 to 2 in the exponent of the denominator. This accounts for the fact that the exponential correlation function is quite a good approximation to the truth for many practical purposes. The difference between the exponents is not always negligible, however. For the (quite typical) conditions of the North Sea experiment described by Megaw,[1] the exponential correlation would yield an expression for the scattering cross-section which, for a given r.m.s. refractive-index fluctuation, was 7 dB lower than that given by eqn. (14) at 0 nautical miles' range, and 10 dB lower at 300 miles.

The simple expression (11) must also be modified when wave numbers of the same order as, or greater than, k_s are considered. A form for the spectral density in this region has been suggested by Heisenberg;[14] this leads[15] to a proportionality of $F_n(k)$ to

$$k^{-5/3}\left[1+\left(\frac{k}{k_s}\right)^4\right]^{-4/3}$$

where

$$k_s = \left(\frac{3c^2}{8}\right)^{1/4}\left(\frac{\varepsilon}{\nu^3}\right)^{1/4}$$

The constant c is found experimentally[16] to be approximately 45, so that $a_s \simeq 5\cdot9\eta$, where $\eta = \nu^{3/4}/\varepsilon^{1/4}$. This suggests the following expression for $F_n(k)$, to apply approximately for all positive values of k:

$$F_n(k) = 0\cdot80k_1^{-1}\left(\frac{k}{k_1}\right)^2\left[1+\left(\frac{k}{k_1}\right)^2\right]^{-11/6}\left[1+\left(\frac{k}{k_s}\right)^4\right]^{-4/3}(17)$$

This extension of eqn. (14) does not appreciably change the normalization factor. Eqn. (17) represents an infinite spectrum, but, in fact, the decrease of the spectral density, $F_n(k)$, with decrease of k below k_1, and particularly with increase of k over k_s, is rapid; few of the results of this paper are much affected by the use of eqn. (17) rather than the simple finite spectrum given by eqn. (11).

(3.4) Numerical Values of the Spectral Parameters

Numerical values of a_1 and a_s in the troposphere are now considered. The chief difficulty is to estimate values of the time

* Throughout the paper the eddy size a, corresponding to a wave number k, is indexed as π/k.

rate of energy dissipation per unit mass, ε. Brunt[17] has given approximate estimates based on the height distribution of wind velocity, and the details are set out in Table 1.

Table 1

VALUES OF THE DISSIPATION RATE, ε

| Height interval | Time rate of energy dissipation in height interval | | ρ | ε |
	Total	Per kilometre		
km	watts/m² =ergs/cm³/sec	MW/km³	g/cm³	cm²/sec³ =ergs/g/sec
0–1	3	3	$1\cdot2\times10^{-3}$	~25
1–10	2	$0\cdot2$	7×10^{-4}	~3
Total, 0–10	5	$0\cdot5$	7×10^{-4}	~7 (average)

Interpolating, the values of ε at 1 and 10 km are approximately 10 and 1 erg/g/sec. Only approximate values will be required in the following.

It will be shown later that astronomical observations suggest the value of about 400 m for a_1. This will be roughly a mean value over the troposphere; probably a_1 increases slowly with height.

It may be noted that the unrealistic value of about 10 cm provisionally adopted by Booker and Gordon[2] for the scale of the refractive-index fluctuations has a much less serious effect on their results than might be expected; their final expression (2v) for the scattering cross-section, regarded as an equation for determining l from experimental results, has two positive real roots of which, in effect, the smaller and less plausible was chosen.

From the known values of the air density, the values of ν, and hence of $a_s \simeq 5\cdot9\nu^{3/4}/\varepsilon^{1/4}$, may be derived. These are given in Table 2, together with estimates of the variable component of velocity $\tilde{u}$ given by:

$$\tilde{u} \simeq \left(\frac{8}{\sqrt3}\frac{\varepsilon a_1}{c\pi}\right)^{1/3} \quad \cdots \quad (18)$$

taking, as before, $c \simeq 0\cdot45$. A wide inertial sub-range is always present in the troposphere.

Table 2

VALUES OF THE SPECTRAL PARAMETERS

		At 1 km	At 10 km
a_1, metres	.. ..	~300	~500
ε, ergs per gramme per second	..	10	1
u, metres per second ..	..	$1\cdot0$	$0\cdot6$
ν, square centimetres per second	..	$0\cdot15$	$0\cdot35$
η, centimetres ..	.. ..	$0\cdot14$	$0\cdot46$
a_s, centimetres ..	.. ..	$0\cdot8$	$2\cdot8$

Originally, Megaw[1] assumed the incorrect value of 10 cm for a_s, but this was corrected subsequently.[11] There is now no material discrepancy with the result $\eta \simeq 0\cdot1$ cm of Batchelor.[18] In the following, the values 400 m and 2 cm of a_1 and a_s will be used throughout.

(4) THE SCATTERING OF LIGHT AND RADIO WAVES

The results which have been derived for the spectrum of the refractive-index fluctuations are now applied to the discussion of the fluctuations produced by scattering in the transmission along free-space paths of light and radio waves, and subsequently to the scatter propagation of radio waves to distances well beyond the radio horizon.

(4.1) Scattering from a 'Fourier Element'

Consider the contribution to the scattered energy from a volume large compared with the scale of the fluctuations, for a small interval δk of the wave-number spectrum over which the fluctuations extend. The volume element, a^3, corresponding to the value of k representative of this interval, will be called a 'Fourier element'. The total contribution to the scattered energy for the interval δk is given by the sum of the contributions from all the Fourier elements contained in the volume under consideration. In an element selected at random within this volume, all combinations of relative phasing of the Fourier components of N in the three co-ordinates will be equally probable. The r.m.s. scattered field from the whole volume is obtained as the resultant of randomly-phased contributions from all the Fourier elements. In this process, the field from a single element is evaluated for the co-phased variation of N in the three co-ordinates which results in the maximum scattering, i.e. maximum N at the centre of the element.

Omitting the sinusoidal time variation, let E_i be the field strength incident on a Fourier element from a source at a distance r_1 sufficiently large for E_i to be regarded as constant over the element. Since the magnitude of the refractive-index fluctuation is very small, the modification of field strength within the element arising on this account may be neglected. The scattered field at a distance r_2 from the element is required.

Cartesian co-ordinates are chosen with the origin at the centre of the element, the x-axis being in the direction of propagation, and the y-axis in the direction of the electric vector. Then, if ω is the angular frequency of the radiation, the displacement current density in the y-direction arising from the fluctuation N at any point in the element is, since $N \ll 1$, given by

$$J_y = 2N\epsilon_0 \omega E_i \quad \cdots \quad (19)$$

In terms of the radiation wavelength λ, eqn. (19) becomes

$$\mathbf{J}_y = E_i N / 30\lambda \quad \cdots \quad (20)$$

The total current over the central cross-section of the element perpendicular to the electric vector ($y = 0$) is therefore

$$J = \frac{E_i}{30\lambda} \iint N dx dz \quad \cdots \quad (21)$$

Writing $\quad N = N_k(\sqrt{2}) \cos\frac{\pi x}{a} \cos\frac{\pi y}{a} \cos\frac{\pi z}{a} \quad \cdots \quad (22)$

for the co-phased distribution of N over the element (i.e. over the interval $\pm a/2$ in each co-ordinate), where N_k is the r.m.s. refractive-index fluctuation in the spectral interval dk, the following results after integration:

$$J = \frac{(\sqrt{2})E_i}{30\lambda} \left(\frac{2}{\pi}\right)^2 a^2 N_k \quad \cdots \quad (23)$$

where $N_k = \tilde{N}[F_n(k)dk]^{1/2}$.

At a distance $r_2 \gg a^2/\lambda$, in the direction of propagation, the scattered field due to current in any part of the element a^3 has the same phase, and therefore amounts to

$$E_{s0} = 30J\frac{2\pi a}{\lambda r_2}\frac{2}{\pi} \quad \cdots \quad (24)$$

for the whole element, including the effect of the variation of N in the y-co-ordinate. (The restriction $r_2 > 2a^2/\lambda$, corresponding to a phase difference $< \pi/8$ between the contributions from current elements separated by a distance $a/2$, is sufficient for practical purposes.)

Combination of eqns. (23) and (24) gives

$$E_{s0} = \sqrt{2}\left(\frac{4}{\pi}\right)^2 \frac{E_i a^3}{\lambda^2 r_2} N_k \quad \cdots \quad (2\ )$$

If, however, $a \ll \lambda$, the field due to current in any part of the element has the same phase, irrespective of the direction of observation. The element behaves as a dipole radiating equally in all directions in the zx-plane. In any other direction, making an angle β with that of the incident electric polarization, the scattered field is reduced by the dipole factor $\sin \beta$, and is given everywhere (except in the immediate neighbourhood of the element) by

$$E_s = \sqrt{2}\left(\frac{4}{\pi}\right)^2 \frac{E_i a^3}{\lambda^2 r_2} \sin \beta N_k \quad \cdots \quad (2\ )$$

subject only to $a \ll \lambda$. This is the Rayleigh scattering law for a small dielectric element of volume a^3, with a numerical coefficient appropriate to the distribution of N.

If a/λ is comparable with, or greater than, unity, the scattered field in directions off the axis of propagation is further reduced by destructive interference between the fields of currents in different parts of the element. For a point of observation whose direction from the scattering element makes a small angle (scattering angle) with the axis of propagation, the Fourier transform of the cosine amplitude distribution gives

$$E_s = \frac{\cos\left(\frac{\pi}{2}\sin\gamma\Big/\frac{\lambda}{2a}\right)}{1 - \left(\sin\gamma\Big/\frac{\lambda}{2a}\right)^2} E_{s0} \sin\beta \quad \cdots \quad (2\ )$$

Here the polar factor preceding E_{s0} corresponds to a main lobe of semi-angle about $\lambda/2a$, with monotonically decreasing side lobes for $\gamma \gg \lambda/2a$. Subject to the latter condition, the relative magnitudes of the side-lobe maxima are represented by the factor $[\sin\gamma/(\lambda/2a)]^{-2}$, and the contribution to the r.m.s. scattered field is given, to a good approximation, by inserting $1/\sqrt{2}$ times this factor in eqn. (25). Thus eqn. (27) becomes

$$E_s = \left(\frac{2}{\pi}\right)^2 \frac{E_i a}{r_2 \sin^2 \gamma} \sin \beta N_k \quad \cdots \quad (2\ )$$

Let L be the distance between source and point of observation and E_0 the free-space field strength produced by the source at this distance. When the scattering angle γ is small, as assumed in the remainder of the paper, L may be identified with the path length ($r_1 + r_2$) and $\sin \beta$ taken equal to unity. Then, assuming that no incident energy arrives by indirect paths, $E_i = E_0 L/r_1$, and from eqns. (25) and (28), writing γ for $\sin \gamma$ and unity for $\sin \beta$, is obtained

$$\frac{E_s}{E_0} \simeq \sqrt{2}\left(\frac{4}{\pi}\right)^2 \frac{L}{r_1 r_2} \frac{a^3}{\lambda^2} N_k \quad \cdots \quad (2\ )$$

provided that $\quad \gamma < \frac{\lambda}{2a}$, and $r_2 > \frac{2a^2}{\lambda}$

and $\quad \frac{E_s}{E_0} \simeq \left(\frac{2}{\pi}\right)^2 \frac{L}{r_1 r_2} \frac{a}{\gamma^2} N_k \quad \cdots \quad (3\ )$

provided that $\quad \gamma > \frac{\lambda}{2a}$

These results are often, though not always, applicable to the scattering of the shorter radio waves, of the order of centimetres to metres in wavelength. In optical terms, eqn. (27) corresponds to Fraunhofer diffraction of the incident field by a Fourier element. The value $2a^2/\lambda$ of r_2 may be taken as the boundary

etween Fraunhofer and Fresnel diffraction and the point which the pattern begins to show a resemblance to the eometrical form of the scattering element. For smaller stances it becomes increasingly difficult to provide an ade-uate approximation to the pattern in electromagnetic terms. nally, when $r_2 \ll 2a^2/\lambda$ (as is often the case for light waves), e pattern corresponds closely to the geometrical form of e scattering element, provided that $a \gg \lambda$. The side-lobe attern beyond the geometrical boundary, corresponding to qn. (30), is then completely insignificant in magnitude. In this se it is only the fluctuations in the direction of propagation hich are of practical interest and, following Scott,[19] a simple lculation in terms of geometrical optics is employed. Recent iticisms of this assumption of validity of geometrical optics r $r_2 \ll a^2/\lambda$, $a \gg \lambda$, by Fellgett[20] have been replied to by egaw.[21]

The effect of the same distribution of N on para-axial rays cident on a Fourier element is now considered. Owing to the nallness of N, the linear deviation in traversing the element is nall compared with a, so that the transverse gradient of N ay be considered constant within the element for any ray, nd the y- and z-deflections independent of one another. Thus e small angular deviation, Γ, say, in traversing the element is iven in component form by

$$\Gamma_y \equiv \frac{dy}{dz} = -2(\sqrt{2}) \sin \frac{\pi y}{a} \cos \frac{\pi z}{a} N_k$$

$$\Gamma_z \equiv \frac{dz}{dx} = -2(\sqrt{2}) \cos \frac{\pi y}{a} \sin \frac{\pi z}{a} N_k$$

hich follow on writing down the expressions for d^2y/dx^2 dN/dy, $d^2z/dx^2 = dN/dz$, and integrating over the range from a/2 to $+a/2$. Hence the linear deviation at distance r_2 is iven by

$$\Delta_y = -2(\sqrt{2})r_2 \sin \frac{\pi y}{a} \cos \frac{\pi z}{a} N_k$$

$$\Delta_z = -2(\sqrt{2})r_2 \cos \frac{\pi y}{a} \sin \frac{\pi z}{a} N_k$$

or rays initially separated in the y-co-ordinate by a small stance δy, the difference in linear deviation at distance $(r_2 \ll 2a^2/\lambda)$ is given by $\delta y \, dy/dy(\Delta_y)$. The relative change of paration is therefore

$$-2\pi(\sqrt{2})\frac{r_2}{a} \cos \frac{\pi y}{a} \cos \frac{\pi z}{a} N_k$$

nd, since the same result follows in the z co-ordinate, the lative change in field strength on the axis of propagation at stance r_2 is given by the negative value of this expression.

The average effect of the passage of a pencil of rays through a ngle Fourier element is obtained by taking mean values over $-a/2 < y < a/2, -a/2 < z < a/2$). Thus the average magnitude f the angular deviation in one co-ordinate is given by

$$\Gamma = \frac{8\sqrt{2}}{\pi^2} N_k \quad . \quad . \quad . \quad . \quad (31)$$

nd correspondingly for the linear deviation,

$$\Delta = \frac{8\sqrt{2}}{\pi^2} r_2 N_k . \quad . \quad . \quad . \quad . \quad (32)$$

milarly, the average magnitude of the relative change in field re ngth is given by

$$\frac{E_s}{E_0} = \frac{8\sqrt{2}}{\pi} \frac{r_2}{a} N_k \quad . \quad . \quad . \quad . \quad (33)$$

As noted above, these expressions apply when $a \gg \lambda$ and $r_2 \ll 2a^2/\lambda$. As r_2 or λ is increased, the pattern produced by the element is increasingly blurred by diffraction, and the relative fluctuation of field strength on the axis of propagation is reduced eventually to the value given by eqn. (29) instead of eqn. (33).

(4.2) Integration over Space

The integral effect of the Fourier elements in a volume large compared with a^3 is now considered. For the whole volume the mean-square relative field-strength fluctuation is

$$\left(\frac{\tilde{E}}{E_0}\right)_k^2 = \int_v \left(\frac{E_s}{E_0}\right)^2 \frac{dv}{a^3} . \quad . \quad . \quad . \quad (34)$$

where E_0, as before, is the free-space field produced by the source at the distance of the point of observation, and E_s is the scattered field for a single element a^3. In general, E_s depends on the position of an element within v and on the direction of the point of observation from the element. The case of Rayleigh scattering is the simplest: eqn. (29) applies, with $a \ll \lambda$ replacing the previous restrictions, and substitution in eqn. (34) gives

$$\left(\frac{\tilde{E}}{E_0}\right)_k^2 = \left[\sqrt{2}\left(\frac{4}{\pi}\right)^2 \frac{La^{3/2}}{\lambda^2} N_k\right]^2 \int_v \frac{dv}{(r_1 r_2)^2} \cdots \quad . \quad (35)$$

which can readily be integrated if the volume v is specified in a simple manner. The same result holds for any value of a/λ, but for large values it does so only for scattering angles appreciably less than $\lambda/2a$ at all points within v, and for distances greater than $2a^2/\lambda$. From eqn. (30) the corresponding result for scattering angles larger than $\lambda/2a$ is

$$\left(\frac{\tilde{E}}{E_0}\right)_k^2 = \left[\left(\frac{2}{\pi}\right)^2 \frac{L}{a^{1/2}} N_k\right]^2 \int_v \frac{dv}{\gamma^4 (r_1 r_2)^2} \cdots \quad . \quad (36)$$

Here the integration is complicated by the variation of γ within v.

The conditions for the use of geometrical optics are $a \gg \lambda$ and $r_2 \ll L \ll L_k$; here, for any eddy wave-number k, L_k denotes the transition distance $2a^2/\lambda$, and L denotes the path length for a wholly tropospheric path, and for a stellar source the distance from the receiver of the most distant element contributing significantly to the scintillation. Then

$$\left(\frac{\tilde{E}}{E_0}\right)_k^2 = \sum_{\delta r} \left(\frac{8\sqrt{2}}{\pi}\right)^2 \left(\frac{N_k}{a}\right)^2 r^2 \delta r \quad . \quad . \quad . \quad (37)$$

the summation being made along the path of the light beam; δr is a length element along the path, large compared with a but sufficiently small compared with L for any variation of (N_k/a) over δr to be neglected. Then

$$\left(\frac{\tilde{E}}{E_0}\right)_k = \frac{8\sqrt{2}}{\pi} a^{-1/2} \sqrt{\left[\int_0^L \left(\frac{N_k}{a}\right)^2 r^2 dr\right]} \quad . \quad . \quad (38)$$

$$\tilde{\Gamma}_k = \frac{8\sqrt{2}}{\pi^2} a^{1/2} \sqrt{\left[\int_0^L \left(\frac{N_k}{a}\right)^2 dr\right]} \quad . \quad . \quad . \quad (39)$$

$$\tilde{\Delta}_k = \frac{8\sqrt{2}}{\pi^2} a^{1/2} \sqrt{\left[\int_0^L \left(\frac{N_k}{a}\right)^2 r^2 dr\right]} \quad . \quad . \quad (40)$$

The linear deviation will not require further consideration; expressions for $\tilde{\Delta}$, corresponding to the results to be derived for $(\tilde{E}/E_0)$ and $\tilde{\Gamma}$, can be obtained in a similar way.
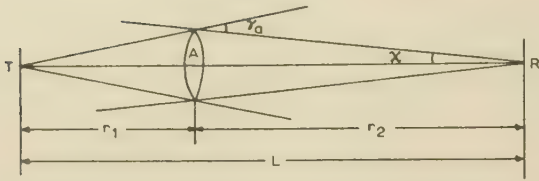
Fig. 1.—Geometry for line-of-sight case (diffraction).

The volume integral (35) applicable to the case $L \gg L_k$ (diffraction) is readily evaluated for line-of-sight propagation (see Fig. 1).

The contributions from the side lobes of elements off the axis may be neglected in comparison with those from the main lobes of elements on or near the axis. The scattered field from each element is accordingly taken equal to E_{s0} over the half-energy beamwidth, which is $\gamma_a = \lambda/2a$, and equal to zero for $\gamma > \gamma_a$. The angle χ is equal to $(L - r_2)\gamma_a/L$, so that the cross-sectional area is $\pi(r_2\chi)^2 = \pi[\gamma_a r_2(L - r_2)/L]^2$; the volume element dv in eqn. (35) is $\pi(r_2\chi)^2 dr_2$. The form of the integrand of eqn. (35) is valid only when both $r_1 > L_k$ and $r_2 > L_k$; it can, however, be shown that the contributions to $(\tilde{E}/E_0)_k^2$ from elements for which $r_1 < L_k$ or $r_2 < L_k$ are negligible when $L \gg L_k$. Then eqn. (35) gives

$$\left(\frac{\tilde{E}}{E_0}\right)_k = \frac{8\sqrt{2}}{\pi^{3/2}} \lambda^{-1} L^{1/2} a^{3/2} \frac{N_k}{a} \qquad . \quad . \quad . \quad (41)$$

The evaluation of the integrals (35) and (36) in the case of beyond-horizon radio propagation is considered later.

For path lengths and eddy sizes such that ray theory is valid, the expressions for $(\tilde{E}/E_0)_k$ and $\tilde{\Gamma}_k$ for a line-of-sight path with a tropospheric source may be obtained by taking N_k/a independent of r in eqns. (38) and (39). This follows since the variations of the intensity of the refractive-index fluctuations along the path, due to changes of height, are small. These require consideration for stellar scintillation, which is discussed in the next Section.

(4.3) Application to Stellar Scintillation

(4.3.1) Integration over the Height Range.

In view of the smallness of Δn at great heights, little error results from neglecting refractive-index fluctuations above a height H, somewhat greater than that of the tropopause. Then, for zenith distances ζ not too near $90°$, the path length $L \simeq H \sec \zeta$ (it can be shown that the effect of mean refraction on the path length is negligible). For most of the range of eddy sizes, $L \ll L_k$, and hence $r = r_2 \ll L_k$; also, the condition $a \gg \lambda$ is always satisfied for light waves, so that geometrical optics may be used. The small-eddy end of the spectrum is considered later.

It is assumed that, for any eddy size at all heights h up to H, N_k/a is proportional to the time-averaged gradient, $\bar{n}' (= dn/dh)$ of refractive index. Otherwise the spectrum and its parameters a_1 and a_s are assumed to be independent of height; this assumption is seriously in error only within about 100 m of the ground, and the contribution to the total scintillation from this small height range is not significant.

The ratio $r/L \simeq h/H$ will be denoted by ξ, and the ratio $[\bar{n}'(h)/\bar{n}'(0)]$ by $G(\xi)$. Then, for all values of ξ, $N_k/a = (N_k/a)_s G(\xi)$, the suffix s denoting the value at sea-level. Eqns. (38) and (39) then give

$$\left(\frac{\tilde{E}}{E_0}\right)_k = \frac{8\sqrt{2}}{\pi} a^{-1/2} \left(\frac{N_k}{a}\right)_s (\sec \zeta)^{3/2} H^{3/2} \sqrt{\left[\int_0^1 G^2(\xi)\xi^2 d\xi\right]} \quad (42)$$

$$\tilde{\Gamma}_k = \frac{8\sqrt{2}}{\pi^2} a^{1/2} \left(\frac{N_k}{a}\right)_s (\sec \zeta)^{1/2} H^{1/2} \sqrt{\left[\int_0^1 G^2(\xi) d\xi\right]} \quad (4$$

In the remainder of this Section (4.3) the values of N_k appropria to sea-level are used and the suffix s is omitted.

The values of $\bar{n}'(h)$ adopted are given in Table 3: they a average values, representative of temperature latitudes in a seasons, based on radiosonde observations (of temperature at humidity) at heights up to 11 km; at greater heights a consta temperature of $217°$ K is assumed. The values for radio wav are the greater in the lower troposphere because of the great refractive index of water vapour at radio frequencies. In t lowest 2 km, large departures from the average values given a common. These may have a considerable effect on free-spa radio field-strength fluctuations and beyond-horizon rad scatter propagation. However, their effect on stellar scintillatio especially for $(\tilde{E}/E_0)_k$ [see eqn. (42)], is not of great importanc and will tend to cancel if a mean of observations at differe times and places is taken.

Table 3

MEAN REFRACTIVE INDEX GRADIENTS

h km					For light m^{-1} $\times 10^{-8}$	For radio m^{-1} $\times 10^{-8}$
0	..	..	..	..	2·9	5·0
0·5	..	..	..	..	2·8	4·7
1	..	..	..	..	2·7	4·3
1·5	..	..	..	..	2·5	4·0
2	..	..	..	..	2·4	3·5
3	..	..	..	..	2·2	2·9
4	..	..	..	..	2·0	2·4
5	..	..	..	..	1·9	2·1
6	..	..	..	..	1·8	1·9
8	..	..	..	..	1·5	1·5
10	..	..	..	..	1·3	1·3
12	..	..	..	..	1·1	1·1
14	..	..	..	..	0·9	0·9
16	..	..	..	..	0·7	0·7
18	..	..	..	..	0·5	0·5

In the following, H will be taken as 15 km; numerical int gration gives (for light waves):

$$\int_0^1 G^2(\xi)\xi^2 d\xi = 0·057 \quad . \quad . \quad . \quad (4$$

and

$$\int_0^1 G^2(\xi) d\xi = 0·36 \quad . \quad . \quad . \quad . \quad (4$$

(For other values of H, between 10 and 20 km, the first integr is approximately proportional to $H^{-3/2}$ and the second to $H^{-3/4}$

(4.3.2) Integration over the Spectrum.

Since the coefficients N_k are assumed to be uncorrelate integration over the spectrum of eddy sizes is performed by int grating the squares $a^{-1}(N_k/a)^2$ and $a(N_k/a)^2$ with respect to (actually k/k_1, to keep the dimensions correct). The 'basi spectrum is taken as eqn. (11), assumed to be applicable fo $k_1 \leq k \leq k_s$; allowance is made as follows for some modifica tions of eqn. (11) which are necessary in practice. One or mor additional cut-offs of the spectrum could be catered for merel by changing the limits of integration to the cut-off wave number Some other effects may be taken into account by multiplyin the spectral density (11) by an appropriate factor. The int grating effect of the finite aperture, A, of the telescope gives factor $\mu_3^2(k)$, equal to (the square of) the spectrum of the apertur This is regarded as being one-dimensional, since it is difficu to take account of the smoothing effect of its second dimensio

e. perpendicular to the mean wind; this, however, is probably not important. $\mu_3^2(k)$ is given by

$$\mu_3^2(k) = \left(\frac{Ak}{\sqrt{2}}\right)^{-2} \quad \text{for } k > \frac{\sqrt{2}}{A} \quad . \quad . \quad . \quad (46)$$

and

$$\mu_3^2(k) = \left[1 - \left(\frac{Ak}{\sqrt{2}}\right)^2\right]^2 \quad \text{for } k < \frac{\sqrt{2}}{A} \quad . \quad . \quad (47)$$

the full integrating effect of the aperture, which applies for eddy sizes $a = \pi/k$ appreciably smaller than $A\pi/\sqrt{2}$, is thus to reduce the index of the basic spectral law of r.m.s. fluctuation by unity. A cut-off of the higher wave numbers by the instrumental time-constant can also be taken into account: the instrumental response may be represented approximately by a simple exponential with time-constant τ, and then the spectral density derived from the records is $\mu_2(k)$ times the true spectral density, where

$$\mu_2(k) = [1 + (k\tau U)^2]^{-1/2} \quad . \quad . \quad . \quad . \quad (48)$$

Here, eddy frequencies, f, (which throughout will be measured in cycles (not radians) per second, so that the eddy period t corresponding to f is f^{-1}) have been converted to wave numbers, by the relation $f = (Uk/2\pi)$, where U is a constant. This procedure will be further discussed presently. For wave numbers not satisfying $k_1 \ll k \ll k_s$ the extension (17) of the Kolmogoroff spectrum is required, giving another multiplying factor

$$\mu_1^2(k) = \left[1 + \left(\frac{k}{k_s}\right)^4\right]^{-4/3} \left(\frac{k}{k_1}\right)^{11/3} \left[1 + \left(\frac{k}{k_1}\right)^2\right]^{-11/6} \quad (49)$$

This last factor must also be taken into account in estimating the r.m.s. value, which will be denoted by $\tilde{N}'$, of N/a. Including these factors μ_1^2, μ_2 and μ_3^2, eqns. (42) and (43) become

$$= 1 \cdot 57 \times 10^6 (\sec \zeta)^{3/2} \tilde{N}' k_1^{1/2} \pi^{-1/2}$$

$$\sqrt{\frac{\int_0^\infty \left(\frac{k}{k_1}\right)^{4/3} \mu_1^2(k) \mu_2(k) \mu_3^2(k) d\left(\frac{k}{k_1}\right)}{\int_0^\infty \left(\frac{k}{k_1}\right)^{1/3} \mu_1^2(k) d\left(\frac{k}{k_1}\right)}} \quad (50)$$

$$= 1 \cdot 73 \times 10^7 (\sec \zeta)^{1/2} \tilde{N}' k_1^{-1/2} \pi^{1/2}$$

$$\sqrt{\frac{\int_0^\infty \left(\frac{k}{k_1}\right)^{-2/3} \mu_1^2 \mu_2 \mu_3^2 d\left(\frac{k}{k_1}\right)}{\int_0^\infty \left(\frac{k}{k_1}\right)^{1/3} \mu_1^2(k) d\left(\frac{k}{k_1}\right)}} \quad \text{seconds of arc} \quad (51)$$

It is more convenient to work in terms of $\tilde{N}'$, rather than the r.m.s. fluctuation $\tilde{N}$ of refractive index itself; the relation connecting $\tilde{N}'$ and $\tilde{N}$ is given in eqn. (74). It is uniquely defined by the spectrum. For the simple spectrum (11) the spectral densities $(\tilde{E}/E_0)_k^2$ and $\tilde{\Gamma}_k^2$ are proportional respectively to $k^{4/3}$ and $k^{-2/3}$.

3) Deductions from the Observations.

The available observations of stellar brightness and angular-deflection fluctuations have all been considered in detail, and allowance made in the above manner for aperture integration, instrumental cut-off and the extended form of the spectrum, as necessary. Only the results will be discussed here. The observations are of variations in time. These are related to space

variations by making the assumption that the atmosphere behaves as if it were a homogeneously turbulent stream with a uniform mean velocity (denoted by U), large in comparison with the turbulent velocities, across the line of sight. Then $k = 2\pi f/U$. In practice, in the troposphere there are usually appreciable variations of the horizontal mean wind velocity with height (in addition to the rapid variation near the ground due to surface friction, which can be neglected since it affects only a small height range, contributing little to the total scintillation). It will be shown, however, that different sets of observations, interpreted on the assumption of uniform horizontal mean wind, yield nearly the same form (actually the Kolmogoroff form) for the refractive-index spectrum, although the mean-wind profiles must vary considerably on the different occasions concerned. No really satisfactory explanation of this unimportance of the mean-wind profile has yet been produced.

A plot of available angular deflection spectra was given in a previous paper;[1] the spectral densities (which are of r.m.s. fluctuation) have all been 'corrected' to zero zenith distance by multiplication by the factor $(\cos \zeta)^{1/2}$ indicated by eqn. (51). The plot shows that for periods t ($= f^{-1}$) up to about 50 sec the spectral density is not inconsistent with the law of variation $t^{1/3}$, which, taking $k = 2\pi f/U$, corresponds to the expected variation as $k^{-1/3}$ in the inertial sub-range. The records available do not extend to small eddy sizes (high frequencies), so that ray theory is valid and aperture integration is negligible. The absence of the high-frequency components means that the true value of $\tilde{\Gamma}$ cannot be obtained from the records. It is now estimated from the spectrum, which is assumed to have the Kolmogoroff form (11) in the inertial sub-range, and, for all wave numbers, to be given by eqn. (17); the consistency of the values of $\tilde{N}'$ to be derived from the angular deflection and the brightness observations will provide a test of the correctness of this assumption. The value of $\tilde{N}'$ obtained from the deflection results will be a mean over the four sets of observations plotted. It is necessary to assume a numerical value for the mean horizontal wind velocity U, and the value of 16 m/sec, representing a mean over the troposphere, is taken. Taking the above figure of 50 sec as the period t_1 corresponding to the eddy size a_1, this value of U gives $a_1 \simeq 400$ m. The frequency $f_s = t_s^{-1}$, corresponding to the known value 2 cm of a_s, is 400 c/s. $\tilde{\Gamma}^2$ is then the integral over frequency (actually over f/f_1 for dimensional reasons) of the expression (17) (with $k = 2\pi f/U$) multiplied by a constant of proportionality, obtainable from the graph. The result for the true value of $\tilde{\Gamma}$ is $1 \cdot 14$ seconds of arc. This is proportional to $k_s^{1/6}$, and thus differs from the value given in Table 2 of Reference 1, since a_s was there taken as 10 cm. In deriving $\tilde{N}'$ from this by means of eqn. (51), $\mu_3^2(k)$ and $\mu_2(k)$ are, of course, put equal to unity. Numerical evaluation of the integrals in eqn. (51) shows that the only change produced by the use of eqn. (17) instead of eqn. (11) is to reduce the numerical coefficient by 2%, so that

$$\tilde{\Gamma} = 1 \cdot 70 \times 10^7 \pi^{1/2} \tilde{N}' \sqrt{\frac{4}{k_s}} \quad . \quad . \quad . \quad (52)$$

The value of $\tilde{N}'$ obtained is $2 \cdot 4 \times 10^{-7}$ m^{-1}. This differs slightly from the value given in Table 2 of Reference 1 because of its proportionality to $k_s^{2/3}$ and also because of the revision of the definition of N_k [by the introduction of the factor $\sqrt{2}$ in eqn. (22)] since that paper was written.

The spectral density of stellar-brightness fluctuation, deduced from some records of Mikesell et al.,[22] was plotted in Fig. 6 of Reference 11. The information available relates principally to the small-eddy end of the spectrum. If ray theory were valid for all eddy sizes, eqns. (46) and (49) show that, for wave numbers

k appreciably smaller than k_s, but large compared with the wave number, π/A, corresponding to an eddy of size equal to that of the aperture (so that aperture integration is fully effective), the spectral density of r.m.s. brightness fluctuation would be proportional to $k^{-1/3}$. Also, for wave numbers larger than k_s and π/A it would be proportional to k^{-3}. The spectral plot shows a region closely corresponding to the minus-one-third power law. For wave numbers greater than about $\frac{1}{2}k_s$, the spectral density falls off more rapidly with increase of k than ray theory would suggest, owing to the increasing effect of diffraction. The zenith distance of the star observed was $51°$; contributions from refractive-index fluctuations above about 15 km may be neglected, so that the path length L is about 25 km. For the main contribution, L is probably not greater than 10 km. Then the eddy sizes [equal to about $3(\lambda L/2)^{1/2}$] at which the attenuating effect of diffraction becomes appreciable are about 15 and 24 cm respectively, corresponding to $k/k_s = 1/7\cdot5$ and $1/12$, for these two path-lengths. The discrepancy between these values and the value $\frac{1}{2}$ of k/k_s obtained from the plot—which, if true, would correspond to a path length of about 700 m and height of 400 m—probably indicates that the increase of slope in the plot is due principally to diffraction rather than to viscosity, and that the designation of a wave number in the centre of the region of changing slope as k_s is not valid. Values of $\widetilde{N}'$ are now estimated from eqn. (50); the best procedure here is to use the values of $\widetilde{E}/E_0$ estimated from the available records. Thus, the effects of aperture integration and instrumental cut-off have now to be taken into account. The integrals in eqn. (50) are evaluated numerically, the results being the values given in Table 1 of Reference 1, divided by the factor $\sqrt{2}$ arising from the revised definition of eqn. (22). The revision of the adopted value of a_s has negligible effect. These values of $\widetilde{N}'$ vary between $0\cdot3$ and $0\cdot7 \times 10^{-7}\,\mathrm{m}^{-1}$, their mean being $0\cdot5 \times 10^{-7}\,\mathrm{m}^{-1}$.

The observations have also been investigated on the assumption of the spectrum (15) given by an exponential correlation function; a sharp cut-off is inserted at $k = k_s$, and the wave number k_1 may be taken as l^{-1}. In the inertial sub-range $k_1 \ll k \ll k_s$, the spectral density of $\widetilde{\Gamma}$ should now be proportional to $t^{1/2}$ (i.e. to $f^{-1/2}$ or $k^{-1/2}$); a mean slope of $+\frac{1}{2}$ is not excluded in the plot given in Reference 1. On the assumption that the spectrum of $\widetilde{\Gamma}$ derived from eqn. (15) fits the observations for all wave numbers, a value of $\widetilde{N}'$ may be deduced in the same way as for the refractive index spectrum (17); numerical integration gives (to within a few per cent) the same result as is obtained by assuming a simple refractive-index spectral density proportional to $k^{-1/2}$, corresponding to eqn. (11). Then the following corresponds to eqn. (52):

$$\widetilde{\Gamma} = 1\cdot70 \times 10^7 \pi^{1/2} \widetilde{N}' \sqrt{\frac{\log_\varepsilon \dfrac{k_s}{k_1}}{k_s}} \quad . \quad . \quad . \quad (53)$$

The value of $\widetilde{\Gamma}$ obtained is $0\cdot49$ second of arc, so that

$$\widetilde{N}' = 6\cdot5 \times 10^{-8}\,\mathrm{m}^{-1}$$

Including aperture integration, the 'exponential' spectrum gives a spectral density of r.m.s. brightness fluctuation proportional to $k^{-1/2}$, compared with $k^{-1/3}$ for the Kolmogoroff spectrum. The experimental results plotted in Reference 11 fit the latter spectrum better. The mean of the values of $\widetilde{N}'$ derived from the results for brightness fluctuation summarized in Table 1 of Reference 1 by numerical integration, using eqn. (15), is $2\cdot5 \times 10^{-7}\,\mathrm{m}^{-1}$. There is a similar difference (although in the opposite sense) between the two deduced values of $\widetilde{N}'$ for

the Kolmogoroff spectrum. A difference of this order cann[ot] in the absence of direct measurements of the refractive-ind[ex] fluctuations, e.g. by refractometer, be regarded as being inco[n]sistent with either spectrum, since the two types of observati[on] were made at different times and places. The important eff[ect] of an apparently small change in the form of the spectrum [is,] however, well illustrated by the change by a factor of 19 in t[he] ratio of the deduced values when taken in the same order.

The spectra deduced from each type of observation are sligh[tly] closer to the Kolmogoroff form than the exponential form. [In] view of the limitations and the small number of the scintillati[on] records available for analysis, the observational support for t[he] Kolmogoroff spectrum is as good as could be expected. Recalli[ng] the substantial support for this form of the spectrum from theo[ry,] its use seems amply justified.

Rough calculations for some spectra, which differ m[ore] markedly than the 'exponential' spectrum from the Kolmogor[off] form (but are assumed to have the same cut-offs) give, in ea[ch] case, deduced values of $\widetilde{N}'$ from the two types of observati[on] which are widely different. It may therefore be concluded th[at] these spectra (including that used by Villars and Weisskop[f]) are inappropriate.

(4.3.4) The Dependence on Zenith Distance.

The dependence of $\widetilde{E}/E_0$ on zenith distance is now discusse[d.] The results in this Section will hold for any form of the refracti[ve] index spectrum, provided that it is independent of the horizont[al] co-ordinates, and also of the height, apart from a variation [of] intensity independent of eddy size. If ray theory could be appli[ed] for all eddy sizes, then for each size (and therefore for th[e] integral over the spectrum), $(\widetilde{E}/E_0)_k$ would be proportional [to] $(\sec \zeta)^{3/2}$.

If the diffraction treatment (Rayleigh scattering) were val[id] for all sizes, i.e. if the path length L exceeded the value of $2a^2[/\lambda]$ for the largest significant eddies, $\widetilde{E}/E_0$ would be proportional [to] $(\sec \zeta)^{1/2}$. In practice, ray theory is valid for all but the smalle[st] eddies when the zenith distance is fairly small (furthermore, th[e] effect of the smallest eddies is reduced by aperture integration[),] and wave theory applies for a considerable portion of the rang[e] of eddy sizes when the zenith distance is not far from $90°$. Th[e] above types of dependence on ζ are quite well supported by th[e] observational results plotted in Fig. 7, Reference 23. Th[e] plot shows a trend towards independence of ζ as $\zeta \to 90°$; th[e] failure of the approximation $L \simeq H \sec \zeta$ is probably n[ot] significant here. However, since the fluctuations are by n[o] means small for stars near the horizon, in the limit of larg[e] path-length an approach to a Rayleigh distribution of brightne[ss] fluctuation, independent of L, is to be expected.[21]

(The assumption made throughout the paper, that the fluctua[tions] discussed are small, is otherwise always valid; it effects [a] great simplification of the work.)

The considerable bandwidth of the light radiation, neglecte[d] up to now, also requires consideration. It makes no differenc[e] when ray theory is valid for all wavelengths and eddy size[s] involved, and thus does not appreciably alter any of the resul[ts] of the foregoing discussion of spectra. However for far-fie[ld] conditions ($L > 2a^2/\lambda$, for all values a) the r.m.s. fluctuatio[n] $\widetilde{E}/E_0$ is proportional to λ^{-1}; in the limit the fluctuations i[n] adjacent intervals of the wavelength range are completely unco[r]related. In the intermediate zone there will be approximatel[y] full correlation if $\Delta\lambda$, the difference between the extreme wav[e]lengths present, is less than λ_0, and approximately no correlatio[n] if $\Delta\lambda > \lambda_0$, where λ_0 is some critical wavelength. For th[e] latter, it seems plausible, after Nettelblad,[24] to take the valu[e] a_s^2/L. L will be taken as $(10 \sec \zeta)\,\mathrm{km}$ and a_s as 2 cm, so tha[t] $\lambda_0 = 4 \times 10^6 \cos \zeta\,\mathrm{cm}$. In starlight the range $\Delta\lambda$ is abou[t]

$\times 10^{-5}$ cm, so that, except when a narrow-band optical filter is used, there will be some reduction in $\tilde{E}/E_0$, effectively increasing the monochromatic reduction by diffraction, except when the telescope aperture is appreciably greater than $a_s\sqrt{(\Delta\lambda/\lambda_0)}$, $4\cdot 4(\sec\zeta)^{1/2}$ cm. This is consistent with the result of Hettelblad[24] with $\zeta \simeq 40°$, that for $A = 8\cdot 9$ cm a change of $\Delta\lambda$ from $2\cdot 2$ to $1\cdot 0 \times 10^{-5}$ cm has a just-negligible effect. When $\lambda \gg \lambda_0$ the contributions of successive intervals of λ_0 in the wavelength range will add randomly, and $\tilde{E}/E_0$ will be multiplied by a factor, $m^{-1/2}$ say,

$$\simeq \left(\frac{\Delta\lambda}{\lambda_0}\right)^{-1/2} = \left[\frac{a_s}{(L\Delta\lambda)^{1/2}}\right] = \frac{a_{min}}{(L\Delta\lambda)^{1/2}}$$

if there is a cut-off at an eddy size $a_{min} > a_s$, e.g. due to the aperture A; taking $a_{min} = A$, $m^{-1/2} \simeq A(\sec\zeta)^{-1/2}(\Delta\lambda)^{-1/2}$, so that in the far-field limit when $L > 2a^2/\lambda$ for all values of a, $\tilde{E}/E_0$ would be approximately independent of A and $\sec\zeta$, and proportional to $(\Delta\lambda)^{-1/2}$.

(4.4) Application to Radio Free-Space Propagation

Radio scintillation, i.e. the small fluctuations of electric-field intensity experienced on line-of-sight paths at radio wavelengths, is now considered. For radio waves the much longer wavelength involved means that, over most of the range of eddy sizes (and for all of it for metre waves), wave theory is required; for $\lambda = 3$ m the extreme values of $L_k = 2a^2/\lambda$ are $2a_s^2/\lambda = 2\cdot 7 \times 10^{-2}$ cm and $2a_1^2/\lambda = 108$ km; for $\lambda = 3$ cm they are $2\cdot 7$ cm and $10\,800$ km, so that, for practical path-lengths, ray theory must be used for the largest eddies. This more general case is now considered. An exact analytical treatment of the eddy sizes in the transition zone, i.e. for values of a, and distances r_2 of the Fourier element under consideration from the receiver, such that $2a^2/\lambda$ lies between about r_2 and $10r_2$, would be very difficult. The best that can be done analytically is to select a particular eddy size, a_m, then use the ray-theory expression (38) for all larger eddies and the wave theory expression (41) for all smaller eddies. In eqn. (38) N_k/a is taken to be independent of the position r on the path, since the height variations along the path are small. The value of a for which eqns. (38) and (41) become equal is chosen for a_m, i.e. $a_m = (L\lambda)^{1/2}$. Strictly, the integrations over the path and over the spectrum should be combined, since, for values of a near a_n, different scattering laws may be required for different values of r_2, e.g. for $r_2 \ll L$ and $r_2 \simeq L$. However, it can be shown that neglect of this fact does not cause serious error, provided that a reasonable value of a_m is employed, such as the present value $(L\lambda)^{1/2}$, for which (mean r_2) = $L/2$ = $\frac{1}{2}(2a_m^2/\lambda)$. Eqn. (54) gives the result for $\tilde{E}/E_0$. In the main factor the spectrum (14) is used, the contribution to $\tilde{E}/E_0$ from large wave numbers being very small. In the introductory factor, negligible error is caused by using eqn. (11) instead of eqn. (17).

$$\frac{}{} = \tilde{N}'\frac{2^{3/2}}{3^{1/2}}\left[\left(\frac{k_s}{k_1}\right)^{4/3} - 1\right]^{-1/2}$$

$$\times\left\{\frac{64}{\pi^3}\lambda^{-2}La_1^3\int_{k_m/k_1}^{k_s/k_1}\left(\frac{k}{k_1}\right)\left[1 + \left(\frac{k}{k_1}\right)^2\right]^{-11/6}d\left(\frac{k}{k_1}\right)\right.$$

$$\left. + \frac{64}{3\pi^2}L^3a_1^{-1}\int_0^{k_m/k_1}\left(\frac{k}{k_1}\right)^5\left[1 + \left(\frac{k}{k_1}\right)^2\right]^{-11/6}d\left(\frac{k}{k_1}\right)\right\}^{1/2} \quad (54)$$

If $(k_m/k_1)^2 \gg 1$, the first integral is approximately

$$\frac{3}{5}\left[\left(\frac{k_m}{k_1}\right)^{-5/3} - \left(\frac{k_s}{k_1}\right)^{-5/3}\right]$$

while if $(k_m/k_1)^2 \ll 1$, the second integral is approximately

$$(k_m/k_1)^6(1/6)$$

otherwise numerical integration must be used. If the numerical values of L and λ are given, graphical integration may be employed, using a sketched junction between the refraction and diffraction ranges. It can be shown that the error in the derived value of $\tilde{E}/E_0$ should not then exceed a factor of about $\sqrt{2}$.

This process has been carried out for the particular case $\lambda = 10$ cm, $L = 100$ km. Several sets of observations for path lengths and wavelengths near these values have been published.[25, 26] The values of $\tilde{E}/E_0$ for $\lambda = 10$ cm and $L = 100$ km which correspond to these experimental values may be obtained to sufficient accuracy (the magnitude of the correction in no case exceeds a factor 2) by assuming proportionality to $L^{5/4}\lambda^{-1/4}$, the law of variation in the transition region. For each set of observations the recorder time-constant was appreciably less than 1 sec, and its effect hence unimportant. The value of a_m is 100 m, the usual values, 400 m and 2 cm, of a_1 and a_s being taken.

The experimental values of $\tilde{E}/E_0$ were each about $0\cdot 5$ dB, and the corrected values average $0\cdot 6$ dB. Denoting the values of $\tilde{N}'$ at any height h appropriate to radio and light waves by $\tilde{N}'(h)$ and $\tilde{N}'_L(h)$ respectively, the derived value of $\tilde{N}'(0)$ is 25×10^{-7} m^{-1}. The astronomical deflection and brightness observations respectively give r.m.s. values of N/a of $2\cdot 4$ and $0\cdot 5 \times 10^{-7}$ m^{-1}. Throughout the discussion of stellar scintillation it has been assumed that $\tilde{N}'_L(h)$ is proportional to the gradient $\bar{n}'_L(h)$ of mean refractive index at all levels. Thus, the mean of the values for $\tilde{N}/a$, namely $1\cdot 45 \times 10^{-7}$ m^{-1}, gives that for $\tilde{N}'_L(0)$ in the astronomical case. Making allowance for the greater effect of water vapour in the contribution to refractive index at radio rather than light frequencies [given by the ratio of $\bar{n}'(0)$ to $\bar{n}'_L(0)$], the corresponding value of $\tilde{N}'(0)$ is deduced to be $2\cdot 5 \times 10^{-7}$ m^{-1}, compared with the 25×10^{-7} m^{-1} derived from the radio observations.

The astronomical observations are relatively insensitive to the condition of the lowest layers of the atmosphere. It may thus be concluded that $\tilde{N}'$ at low levels is considerably greater than is suggested by the assumption of proportionality to the mean gradient. Assuming that the enhancement of $\tilde{N}'$ near the surface decreases with height at the same rate as the water vapour contribution to the mean refractive index gradient, the following expression for $\tilde{N}'$ results:

$$\tilde{N}'(h) = 2\cdot 5 \times 10^{-7}\frac{\bar{n}'(h)}{\bar{n}'(0)}\left\{1 + 9\left[\frac{\dfrac{\bar{n}'(h)}{\bar{n}'_L(h)} - 1}{\dfrac{\bar{n}'(0)}{\bar{n}'_L(0)} - 1}\right]\right\} \quad (55)$$

The values of $\bar{n}'$ and $\bar{n}'_L$ are given in Table 3.

(4.5) Application to Radio Scatter Propagation beyond the Horizon

(4.5.1) Integration over the Scattering Volume.

Radio-wave propagation in the region beyond the horizon is now discussed.

The geometry for the general case is illustrated in Fig. 2. The effective radius of the earth for radio propagation, taking account of the height gradient of mean refractive index $\bar{N}$, is denoted by R, and the angle between the tangent planes to the earth from the transmitting and receiving aerials by α. It will be assumed that the beam widths and α are small in comparison with a
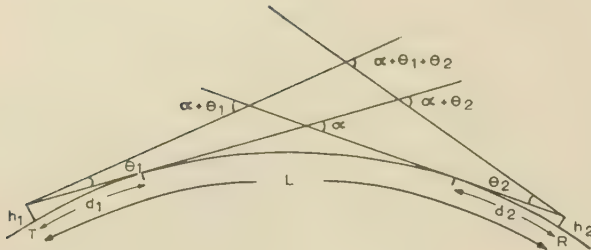
Fig. 2.—Geometry for beyond-horizon case (great-circle plane).

radian. This is usually permissible in practice for centimetre wavelengths; actually, the error is not serious should the beam widths in just one plane not be small. Denoting by d_0 the sum of the horizon distances $\sqrt{(2Rh_1)}$, $\sqrt{(2Rh_2)}$ of the transmitting and receiving aerials, it is assumed that d_0 is sufficiently smaller than L to make the diffracted field negligible.

The heights must still be large enough to keep the elevation of the lowest lobe negligible compared with θ_1 or θ_2. This condition is nearly always fulfilled in practice at centimetre wavelengths, but not necessarily for longer wavelengths. Account of the variation with height of $\widetilde{N}'$, assumed to be given by eqn. (55) and Table 3, is taken by assuming that the value appropriate to a particular height h, whose selection is discussed later, applies for all heights. Then for all values of k, $(N/a)_k$ is assumed to be independent of height.

It is assumed that θ_1 is of the same order of magnitude as θ_2, and ϕ_1 as ϕ_2; this is normally satisfied in practice. It can then be shown that little error is caused by taking the elevation semi-beam-widths both equal to $(\theta_1\theta_2)^{1/2}$, $= \theta_1'$, say, and the azimuth semi-beam-widths both equal to the lesser of ϕ_1 and ϕ_2, $= \phi_1'$, say. The beams are assumed to be 'rectangular'.

First consider the case of off-axis scattering, i.e. $\lambda/2a \ll \gamma \ll 1$. It is then required to evaluate eqn. (34) over the scattering volume v. Then

$$\left(\frac{\widetilde{E}}{E_0}\right)^2_k = 4\iiint \left(\frac{E_s}{E_0}\right)^2_k a^{-3}dv \quad . \quad . \quad . \quad (56)$$

with domain of integration: $0 < \phi < \phi_1'$, $0 < \psi < \dfrac{\pi}{2} - \dfrac{\alpha}{2}$, $0 < \theta < \theta_1$. An element of volume, dv, in these co-ordinates θ, ψ, ϕ takes the form

$$dv = \left(\frac{L}{2}\right)^3 \theta \sec^2\left(\psi - \frac{\alpha}{2} - \theta\right)(1 - C)d\phi d\psi d\theta \quad . \quad (57)$$

Also,

$$r_1 = \frac{L(1 - C)}{2}, \quad r_2 = \frac{L(1 + C)}{2}, \text{ where } C \equiv \theta \tan\left(\psi - \frac{\alpha}{2} - \theta\right)$$

and

$$\gamma = \frac{2\sqrt{[\frac{1}{2}\alpha + \theta)^2 + \phi^2]}}{(1 + C)}$$

Then, substituting the expression (30) for $(E_s/E_0)_k$, consideration of the integral, without further approximation, gives

$$\left(\frac{\widetilde{E}}{E_0}\right)_k = \frac{2^{3/2}}{\pi^2}L^{-1/2}Ra^{1/2}\left(1 - \frac{d_0}{L}\right)^{-1}\left(\frac{N}{a}\right)_k D(p_1, b) \quad (58)$$

$$= a^{1/2}\left(\frac{N}{a}\right)_k Q, \text{ say} \quad . \quad . \quad . \quad . \quad . \quad (59)$$

where $$D^2(p_1, b)$$

$$= \int_0^{p_1} dp\left\{\frac{bp_1}{(1 + p)^2[(1 + p)^2 + bp_1^2]} + \frac{1}{(1 + p)^3}\tan\arctan\left(\frac{bp_1}{1 + p}\right)\right\}$$

$$\left[8\log_\varepsilon\left(1 + \frac{p}{2}\right) - \frac{\left(\frac{p}{2}\right)^2}{\left(1 + \frac{p}{2}\right)^2} - \frac{6\frac{p}{2}}{1 + \frac{p}{2}}\right] \quad (6$$

Here $p = 2\theta/\alpha$ and $b = \phi_1'/\theta_1'$. The factor D gives the magnitude of the well-known feature of scatter transmission of giving a reduction in the effective combined gain of the transmitting and receiving aerials below the sum of their plane-wave values, as a function of the beamwidths. Table 4 gives values of

Table 4

VALUES OF $D(p_1, b)$

p_1	0·1	0·3	1	3	10	30	100
$b = \frac{1}{4}$	0·014	0·059	0·21	0·46	0·72	0·85	0·88
$b = 1$	0·028	0·12	0·37	0·66	0·83	0·88	0·88
$b = 4$	0·055	0·19	0·46	0·72	0·85	0·88	0·89

D, obtained by numerical integration, for representative values of p_1 and b. Interpolation between these will give the values of D for most values of p_1 and b of practical interest.

If for all eddy sizes a, $\lambda/2a < \gamma \ll 1$, $\widetilde{E}/E_0$ is obtained by (numerical) integration of eqn. (59) over the spectrum, assumed to be of the form given in eqn. (17). Then

$$\frac{\widetilde{E}}{E_0} = Qa_1^{1/2}\widetilde{N}'(h)1\cdot 96\left[\frac{\left(\frac{k_s}{k_1}\right)^{1/3} - 1}{\left(\frac{k_s}{k_1}\right)^{4/3} - 1}\right]^{1/2} \quad . \quad . \quad (61$$

[If the simple form eqn. (11), assumed to apply for all values of in the range $k_1 \leqslant k \leqslant k_s$, is used, the only change in eqn. (61) is that the numerical coefficient becomes 2.]

In the troposphere, $(k_s/k_1)^{1/3} \gg 1$, so that

$$\frac{\widetilde{E}}{E_0} = 2Qa_s^{1/2}\widetilde{N}'(h) \quad . \quad . \quad . \quad (62)$$

Throughout, all lengths are to be measured in metres, unless otherwise stated.

For an eddy size such that $\gamma \ll \lambda/2a$, manipulation of the result of substituting eqn. (29) in eqn. (34), with no further approximations, gives

$$\left(\frac{\widetilde{E}}{E_0}\right)^2_k = \left[\sqrt{2}\left(\frac{4}{\pi}\right)^2 L^{1/2}a^{5/2}\lambda^{-2}\left(\frac{N}{a}\right)_k\right]^2 \phi_1\frac{\alpha}{2}$$

$$\times \left[2(\theta_1')^2\left(1 + \frac{2p_1}{3}\right) + p_1\log_\varepsilon(1 + p_1)\right] \quad (63$$

This is the case of Rayleigh scattering, and holds for any value of $\lambda/2a$, provided that, if $\lambda/2a$ is small, $L > 2a^2/\lambda$. In practice, the first term in eqn. (63) can be shown not to exceed 5% of the second, and is usually much less. Hence

$$\left(\frac{\widetilde{E}}{E_0}\right)_k = \frac{16\sqrt{2}}{\pi^2}L^{1/2}a^{5/2}\lambda^{-2}\frac{N_k}{a}[2\theta_1'\phi_1'\log_\varepsilon(1 + p_1)]^{1/2} \quad (64$$

$$= Sa^{5/2}\frac{N_k}{a}, \text{ say} \quad . \quad . \quad . \quad . \quad (65$$

the condition of validity, $\gamma \ll \lambda/2a$, holds over the whole range $k_1 \leqslant k \leqslant k_s$, a sharp cut-off at $k = k_s$ may be assumed, since the integrand is small for large values of k, and then, if instrumental cut-off is negligible in the range $k \leqslant k_s$,

$$\left(\frac{\tilde{E}}{E_0}\right)_k = \sqrt{\frac{4}{11}} Sa_1^{5/2}\tilde{N}'(h)\left(\frac{k_s}{k_1}\right)^{-2/3} \quad . \quad . \quad (66)$$

In deriving eqn. (66), eqn. (11) is used; if eqn. (17) were employed, a large contribution would have come from $k < k_1$, but in practice, even for wavelengths as large as 10 m, usually

$$\gamma \gg \lambda/2a \text{ for } k < k_1.$$

(4.5.2) The Transition between the Two Scattering Functions.

In general, the transition between the two ranges of wave numbers defined by $\gamma \gg \lambda/2a$ and $\gamma \ll \lambda/2a$ falls within the range k_1–k_s. Strictly, the full expression (27) should be used in the integrations over the scattering volume and spectrum, with the appropriate expression for γ in terms of the geometry of the problem inserted in the polar factor of E_s. However, the variation of γ over the scattering volume is bounded by the angles α below, and $(\alpha + 2\theta_1')$ (somewhat more if $\phi_1' \gg \theta_1'$) above, so that for eddy sizes given by

$$a \gg \frac{\lambda}{2}\frac{R}{(L - d_0)} = a\gamma_1, \text{ say,} \quad . \quad . \quad (67)$$

$\gamma \gg \lambda/2a$ throughout the scattering volume and for sizes given by

$$a \ll \frac{\lambda}{2(\alpha + 2\theta_1')} = a\gamma_2, \text{ say} \quad . \quad . \quad . \quad (68)$$

$\gamma \ll \lambda/2a$ throughout the scattering volume. The integration over the spectrum may then be performed graphically, using a sketched transition between the two ranges of eddy sizes.

(4.5.3) The Metre-Wave Case.

The metre-wave case is, in general, more difficult than the centimetre-wave case because of the effect of earth reflection. In general, this must be taken into account in the volume integral, unless the aerial heights R_1 and R_2 are sufficiently large, when the integration can be carried out in the same way as for centi-metre waves. For the largest wavelengths the relation $\gamma < \lambda/2a$ holds for nearly the whole of the common volume and the whole spectrum. For example, consider a range, L, of 400 km and wavelength, λ, of 10 m. Approximate account of the largest eddies, for which $\gamma > \lambda/2a$, is taken by assuming the expression (56) to hold for all eddy sizes considered, with the large eddy cut-off taken at $a_1' = 2 \cdot 5 \ (\lambda/2\gamma_{max})$ instead of a_1. Taking $\theta_1 = 0 \cdot 1$ rad and $\gamma_{max} = \alpha + 2\theta_1' = 0 \cdot 25$ rad, $a_1 = 50$ m and $k_s/k_1' = 2 \cdot 5 \times 10^3$. The height, h, may be taken as the height of the centre of the scattering volume (in any case the height variation of $\tilde{N}'$ is only slow), here roughly 6 km. Then, if $\theta_1 = \phi_1' = 0 \cdot 1$ rad, eqn. (66) gives $\tilde{E}/E_0 = -79$ dB, which is quite close to typical centimetre-wave results for the same distance. The theory shows directly that the wavelength dependence of $\tilde{E}/E_0$ is only very slight in the centimetre-wave band.

(4.5.4) Further Discussion of the Centimetre-Wave Case.

The remainder of this Section treats the centimetre-wave band, in which the effect of earth reflection, to be considered presently, is usually small. The result, eqn. (59), is valid when $a \gg a\gamma_1$, but eqn. (64) should be used when $a \ll a\gamma_2$. The integration over the spectrum (i.e. over wave number) is then performed graphically, using the refractive-index spectrum (17) and a sketched junction between the two above ranges of eddy sizes. Consideration of the results obtained for a number of different

wavelengths (in the centimetre-wave band) and different distances give an interesting result: in each case integration of eqn. (59) just over the range $k_1 \leqslant k \leqslant k\gamma_1$, using the simple refractive-index spectrum (11), gives very nearly the same result. It will therefore be assumed that this procedure is correct. In the next Section a direct theoretical indication of the importance of the wave number $k\gamma_1$ will appear.

For the shortest paths and wavelengths (e.g. 3 cm) and the largest eddies, the condition of validity of wave theory, or $r_2 \simeq L/2 > 2a^2/\lambda$, is not satisfied; e.g. for the shortest likely path-length in practice (about 80 km) and $\lambda = 3$ cm, it is not satisfied for $a = (L\lambda/4)^{1/2} > 25$ m. For values of a greater than about $3(L\lambda/4)^{1/2}$, ray theory is valid. The contribution from such eddies to the scattered field strength at the receiver is zero, since the scattering angle is appreciable [space variations of the time-averaged refractive index $(1 + \tilde{N})$, especially abrupt variations, can, of course, contribute to beyond-horizon propagation]. This reduction in the contribution to $\tilde{E}/E_0$ from eddy sizes greater than $(L\lambda/4)^{1/2}$ may be taken approximately into account by limiting the integration of eqn. (59) to the range $a \leqslant a_2 = 2(L\lambda/4)^{1/2} = (L\lambda)^{1/2}$. The general result is then

$$\frac{\tilde{E}}{E_0} = Q\tilde{N}'(h)M \quad . \quad . \quad . \quad . \quad (69)$$

where, if $a_2 < a_1$,

$$M = 1 \cdot 39a_1^{-1/6}a_s^{2/3}\left[\left(\frac{a_2}{a\gamma_1}\right)^{1/3} - 1\right]^{1/2} \quad . \quad . \quad (70)$$

and if $a_2 > a_1$, a_1 is written for a_2 in eqn. (70). Since a_2 and $a\gamma_1$ are proportional only to $\lambda^{1/2}$ and λ, respectively, the wavelength dependence of $\tilde{E}/E_0$ is only slight.

Calculation, using the method of Domb and Pryce,[27] for typical cases suggests that for centimetre waves the effect of earth loss and earth reflection can be allowed for by the insertion of an additional multiplying factor in eqn. (69), equal to $+2\frac{1}{2}$ dB at 50 nautical miles ($92 \cdot 7$ km) increasing to $+4$ dB at 300 nautical miles (556 km). Atmospheric attenuation gives another factor in eqn. (69), equal to about -6×10^{-3} dB/km $= -1 \cdot 1$ dB per 100 nautical miles. The unimportance of the small eddies means that the effect of the recorder time-constant is negligible.

Consideration of the variation with θ of the integrand of eqn. (56) shows that the value of h to be used in eqn. (69) is given, in the present case of centimetre waves and narrow beams, by

$$h = \frac{L^2}{2}\left[\frac{(1 - d_0/L)^2}{4R} + 8 \times 10^{-7}\theta_1'\right] \quad . \quad . \quad (71)$$

In the last term a quantity with the dimensions of length (expressed in metres) has been absorbed into the numerical factor. The first term is simply the height of the earth-tangent intersection; this is a lower limit, since the contribution to the scattered field from refractive-index fluctuations below the horizon of either transmitter or receiver can be shown to be completely negligible.

(4.6) Comparison with Other Theoretical Work on Radio Scatter Propagation

No detailed consideration of the relation between stellar scintillation and the spectrum of atmospheric turbulence seems to have hitherto appeared in the literature. Several theoretical discussions[1, 3-6, 28] have been given of beyond-horizon radio scatter propagation. In all, as in the present treatment, multiple scattering is neglected; this is valid for the small refractive-index fluctuations involved.[6] The present treatment was designed to take account of all effects important in practice, with particular

reference to the considerable variations which, in fact, occur within the scattering volume. The alternative treatments available largely ignore these variations, with the result that more rigorous basic formulations can be achieved. These treatments arrive at a result equivalent to the following: the 'scattering cross-section', σ, is given (Batchelor,[6] but using the notation of the present paper) by

$$\sigma = \left(\frac{\tilde{E}}{E_0}\right)^2 \frac{r_1^2 r_2^2}{L^2 v} = \frac{4\tilde{N}^2 \sin^2 \beta}{128 \sin^4 \gamma/2} K^2 F_n(K) \quad . \quad . \quad (72)$$

where

$$K \equiv \frac{4\pi}{\lambda} \sin \frac{\gamma}{2}$$

and r_1, r_2 and γ are assumed to be constant over the scattering volume v. The fact that the spectral density enters into eqn. (72) so directly gives further support for the emphasis laid on the direct consideration of spectra, rather than correlation functions, in the present work. A comparison of eqn. (72) with the results of the present paper may be made by omitting the consideration of variations within v, i.e. using eqn. (36) with γ, r_1 and r_2 assumed to be constant. This gives

$$\left(\frac{\tilde{E}}{E_0}\right)_k^2 = \left(\frac{16}{\pi^2} La^{1/2} \frac{N_k}{a}\right)^2 \frac{v}{\gamma^4 r_1^2 r_2^2} \quad . \quad . \quad . \quad (73)$$

For the spectrum (17), numerical integration shows that

$$\tilde{N} = 1 \cdot 30 a_1^{1/3} a_s^{2/3} \tilde{N}' \quad . \quad . \quad . \quad . \quad . \quad (74)$$

For the purposes of the present comparison it may be assumed that $a_2 > a_1$. A typical value of $(a_1/a\gamma_1)^{1/3}$ for centimetre waves is 4, so that $(a_1/a\gamma_1)^{1/3} - 1 = \frac{3}{4}(a_1/a\gamma_1)^{1/3}$. Then, integration of eqn. (73) [assuming spectrum (11)] only over the range $k_1 \leqslant k \leqslant k\gamma_1 \simeq K$, and substitution from eqn. (74), gives the result

$$\sigma = Bk_1^{2/3} \lambda^{-1/3} \gamma^{-11/3} \tilde{N}^2 \quad . \quad . \quad . \quad (75)$$

where

$$B = \frac{3}{4} 2^{16/3} \pi^{-14/3} [(\sqrt{2})/1 \cdot 30]^2 = 0 \cdot 17 \quad . \quad . \quad (76)$$

The length l in Batchelor's[6] eqn. (10) may be taken equal to the integral scale derived from the spectrum (17), i.e. $0 \cdot 24a_1$. The expression (11) may be used for $F_n(k)$, since $k_1 \ll k \ll k_s$; substitution gives

$$\sigma \simeq \frac{(0 \cdot 80)4\tilde{N}^2}{128(\gamma/2)^{11/3}} \left(\frac{4\pi}{\lambda k_1^{-2}}\right)^{1/3} = 0 \cdot 89 k_1^{2/3} \lambda^{-1/3} \gamma^{-11/3} \tilde{N}^2 \quad . \quad (77)$$

since the polarization factor, $\sin^2 \beta$, is, in practice, nearly equal to unity, and γ is small. Two errors appear to be present in Batchelor's eqn. (10): the index of $\sin \gamma/2$ should be 11/3 and not 13/3, and the last factor should be $(4\pi/\lambda l^2)^{1/3}$. Thus, the present work agrees with the rigorous result, apart from a numerical factor of $5 \cdot 2$ ($7 \cdot 2$ dB).

A comparison between the results of the present paper and those of Batchelor [eqn. (72)] for on-axis scattering may be made as follows: Take the limit of eqn. (72) as $\gamma \to 0$, again using the spectrum (17) for F_n and putting $k_1 = \pi/a_1 = 0 \cdot 24\pi/l$. The result is

$$\sigma = 5 \cdot 9 \tilde{N}^2 \lambda^{-1} \sin^2 \beta (2\pi l/\lambda)^3 \quad . \quad . \quad . \quad (78)$$

Numerical integration of eqn. (35) over the spectrum (17) [with $\int dv/(r_1 r_2)^2$ taken equal to $v/(r_1 r_2)^2$] gives

$$\sigma = 3 \cdot 8 \tilde{N}^2 \lambda^{-1} \sin^2 \beta (2\pi l/\lambda)^3 \quad . \quad . \quad . \quad (79)$$

It may be noted that the result $(2v)$ of Booker and Gordon,[2] when $\gamma = 0$, is

$$\sigma = 4\tilde{N}^2 \lambda^{-1} \sin^2 \beta (2\pi l/\lambda)^3 \quad . \quad . \quad . \quad (80)$$

recalling that $\overline{(\Delta\epsilon/\epsilon)^2}$, in Booker and Gordon's notation, is equal to $4\tilde{N}^2$. The numerical coefficients of eqns. (78) and (79)

are very sensitive to the precise value of the ratio l/a_1 adopted. On account of the high negative power of k in eqn. (35), the principal contribution to eqn. (78) comes from wave number $k < k_1$, and the value of the constant in eqn. (78) depends very much on the accuracy of spectrum (17). In view of this uncertainty, the agreement of eqns. (78), (79) and (80) may be considered satisfactory.

For the case of ray theory, eqn. (38), with the integral taken equal to $(N_k/a)^2 L^3$, may be considered. This is now compared with Bergmann's[29] result, which (in the notation of the present paper) is given by

$$\frac{\tilde{E}}{E_0} = \sqrt{\left[\frac{1}{15} L^3 \tilde{N}^2 \int_0^\infty \frac{\partial^3 \rho(r)}{\partial r^3} \frac{dr}{r}\right]} \quad . \quad . \quad . \quad (81)$$

for the refractive-index correlation function $\rho(r)$. This result is most conveniently applied to the Gaussian correlation $\rho(r) = \varepsilon^{-r^2/l^2}$.

For this,

$$\frac{\tilde{E}}{E_0} = \frac{\pi}{\sqrt{30}} L^{3/2} l^{-3/2} \tilde{N} = 0 \cdot 57 L^{3/2} l^{-3/2} \tilde{N} \quad . \quad (82)$$

The result of integrating eqn. (38) over the spectrum, with N_k taken as the spectral density corresponding to $\rho(r)$, is

$$\frac{8}{\pi^2} \sqrt{\frac{2}{3\pi}} L^{3/2} l^{-3/2} \tilde{N} = 0 \cdot 37 L^{3/2} l^{-3/2} N \quad . \quad (83)$$

Again, the agreement may be regarded as satisfactory.

(5) CONCLUSIONS

The identification of the source of refractive index fluctuations in the atmosphere with a permanently existing state of turbulence has provided the rational basis for investigation of the beyond-horizon radio scatter propagation phenomenon. In view of the limitations of available meteorological knowledge, the spectrum of the refractive-index fluctuations, assumed to be the same as that of the velocity fluctuations, has been predicted from the universal equilibrium theory of turbulence. Analysis of stellar scintillation data has provided confirmation for this theoretical form. This, in conjunction with radio (free-space) fluctuation data, has been used to obtain the height variation of the refractive-index fluctuation intensity.

The present treatment of the radio scattering problem includes detailed consideration of the influence of the geometry of the scattering volume on the signal levels in the region well beyond the horizon. By introducing simplifications in this treatment, it is possible to make comparison with the results of alternative but less-comprehensive treatments based on idealized models. Satisfactory agreement on the form of the scattering cross-section is then achieved.

(6) EPILOGUE

The material presented in the paper has been carried to the point reached by Dr. Megaw at the time of his death. It appears that he had not then made a detailed numerical application of his latest results for comparison with practical observations.

In his treatment of the beyond-horizon scatter propagation problem, a comprehensive discussion of the influence of the geometry of the scattering volume on the magnitude of the received field is given. However, the analysis implies a less rigorous form for the scattering cross-section than that recently obtained by Batchelor.[6] As pointed out in Section 4.6, a reduction in the estimate of the beyond-horizon scatter field of the order of 7 dB is thereby caused. In addition, it appears that the integration over the spectrum of turbulence should be extended to the largest eddy size a_1, and the assumption of a cut-off at an eddy-size a_2 ($a_2 < a_1$) ignored. If these modifications are made, the benefit of Dr. Megaw's thorough treat-

ment of the other factors of practical importance is retained, and it is felt that an exceptionally good theoretical treatment for centimetre waves results.

Table 5 compares Dr. Megaw's experimental results[1] at 10 cm with the estimated results based on the original treatment given in the present paper, and also those including the modifications just described.

Table 5

COMPARISON OF OBSERVED AND CALCULATED SCATTER FIELDS AT 10 CM; NORTH SEA EXPERIMENT, 1949

| Range | $\tilde{E}/E_0$ | | |
| | Observed | Calculated | |
		Unmodified theory	Modified theory
Nautical miles	dB	dB	dB
50	−51	−58	−49
100	−58	−65	−56
150	−60	−74	−65
200	−73	−82	−73
250	−86	−92	−84
300	−96	−102	−94

The calculations of scatter field take account of the change in the adopted value of the smallest eddy size made since the comparison of theory and practice was given in Reference 1. It is seen from the Table that the modifications suggested above give a closer agreement with the observed results than the unmodified treatment. As pointed out in Reference 1, the pronounced 'hump' in the observed results near 150 miles is probably due to the increased intensity of turbulence associated with the cloud layer at about 1–2 km. If this feature is ignored, the modified results for the ratio of scattered to free-space fields are within 2 dB of those observed. While this very close numerical agreement between theory and observation may be partly fortuitous, having regard to the limitations of the available data from which the refractive-index fluctuation intensities were derived, the similarity of slope of observed and calculated curves confirms the adequacy of the theory. Recently, the practicability of obtaining adequately representative data for the lower heights in the atmosphere has been realized; it is desirable to extend the scope of this type of data for practical application of Dr. Megaw's analysis.

(7) ACKNOWLEDGMENTS

(8) REFERENCES

(1) MEGAW, E. C. S.: 'Scattering of Electromagnetic Waves by Atmospheric Turbulence', *Nature*, 1950, **166**, p. 1100.

(2) BOOKER, H. G., and GORDON, W. E.: 'Theory of Radio Scattering in the Troposphere', *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 401.

(3) PEKERIS, C. L.: 'Note on Scattering in an Inhomogeneous Medium', *Physical Review*, 1947, **71**, p. 268.

(4) VILLARS, F., and WEISSKOPF, V. F.: 'The Scattering of Electromagnetic Waves by Turbulent Atmospheric Fluctuations', *ibid.*, 1954, **94**, p. 232.

(5) VILLARS, F., and WEISSKOPF, V. F.: 'On the Scattering of Radio Waves by Turbulent Fluctuations of the Atmosphere', *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 1232.

(6) BATCHELOR, G. K.: 'The Scattering of Radio Waves in the Atmosphere by Turbulent Fluctuations in Refractive Index', Research Report No. EE262, 1955, School of Electrical Engineering, Cornell University.

(7) GALLET, R. M.: 'Aerodynamical Mechanisms producing Electronic Density Fluctuations in Turbulent Ionized Layers', *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 1240.

(8) NORTON, K. A., RICE, P., and VOGLER, L. E.: 'The Use of Angular Distance in Estimating Transmission Loss and Fading Range for Propagation through a Turbulent Atmosphere over Irregular Terrain', *ibid.*, p. 1488.

(9) BATCHELOR, G. K.: 'The Theory of Homogeneous Turbulence' (University Press, Cambridge, 1953).

(10) KOVASZNAY, L. S. G., UBEROI, M. S., and CORRSIN, S.: 'The Transformation between One- and Three-Dimensional Power Spectra for an Isotropic Scalar Fluctuation Field', *Physical Review*, 1949, **76**, p. 1263.

(11) MEGAW, E. C. S.: 'Waves and Fluctuations', *Proceedings I.E.E.*, 1953, **100**, Part III, p. 1.

(12) WHEELON, A. D.: 'Note on Scatter Propagation with a Modified Exponential Correlation', *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 1381.

(13) BOOKER, H. G.: 'A Theory of Scattering by Non-Isotropic Irregularities', *Journal of Atmospheric and Terrestrial Physics*, 1956, **8**, p. 204.

(14) HEISENBERG, W.: 'On the Statistical Theory of Turbulence', *Zeitschrift für Physik*, 1948, **124**, p. 628.

(15) BATCHELOR, G. K.: 'Turbulent Motion', *Reports on Progress in Physics* (Physical Society), 1952, **15**, p. 101.

(16) PROUDMAN, I.: 'A Comparison of Heisenberg's Spectrum of Turbulence with Observation', *Proceedings of the Cambridge Philosophical Society*, 1951, **47**, p. 158.

(17) BRUNT, D.: 'Physical and Dynamical Meteorology' (University Press, Cambridge, 1944).

(18) BATCHELOR, G. K.: 'The Application of the Similarity Theory of Turbulence to Atmospheric Diffusion', *Quarterly Journal of the Royal Meteorological Society*, 1950, **76**, p. 133.

(19) SCOTT, J. M. C.: Private Communication, 1944.

(20) FELLGETT, P. B.: 'Stellar Scintillation', *Quarterly Journal of the Royal Meteorological Society*, 1956, **82**, p. 227.

(21) MEGAW, E. C. S.: 'Stellar Scintillations', *ibid.*, p. 231.

(22) MIKESELL, A. H., HOAG, A. A., and HALL, J. S.: 'The Scintillation of Starlight', *Journal of the Optical Society of America*, 1956, **41**, p. 689.

(23) MEGAW, E. C. S.: Contribution to discussion meeting on 'Stellar Scintillation', *Quarterly Journal of the Royal Meteorological Society*, 1954, **80**, p. 248.

(24) NETTELBLAD, F.: 'Studies of Astronomical Scintillation', Meddelande Från Lunds Astronomiska Observatorium, 1953, Ser. II, No. 130.

(25) MEGAW, E. C. S.: 'Experimental Studies of the Propagation of Very Short Radio Waves', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 79.

(26) McPETRIE, J. S., STARNECKI, B., JARKOWSKI, H., and SICINSKI, L.: 'Over-Sea Propagation Trials on Wavelengths of 3 and 9 cm', *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 243.

(27) DOMB, C., and PRYCE, M. H. L.: 'The Calculation of Field Strengths over a Spherical Earth', *Journal I.E.E.*, 1947, **94**, Part III, p. 325.

(28) STARAS, H.: 'Forward Scattering of Radio Waves by Anisotropic Turbulence', *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 1374.

(29) BERGMANN, P. G.: 'Propagation of Radiation in a Medium with Random Inhomogeneities', *Physical Review*, 1946, **70**, p. 486.

AN INSTRUMENT FOR THE ABSOLUTE MEASUREMENT OF LOW-LEVEL MICROWAVE POWER IN THE 3CM BAND

By Professor A. L. CULLEN, Ph.D., B.Sc.(Eng.), Associate Member, and H. A. FRENCH, Ph.D., B.Sc.(Eng.)

SUMMARY

The paper describes an instrument termed the 'vibration wattmeter' which consists of a cavity in which is suspended a thin metal rod mounted on a fine glass spindle with a quartz fibre suspension. The rod is free to execute torsional oscillations in the horizontal plane.

The oscillating system receives periodic impulses from the inter-action between the metal rod and the electromagnetic field within the cavity, and by the automatic switching of the microwave source a steady amplitude of mechanical oscillations can be maintained. The amplitude of these oscillations is measured by means of a small mirror attached to the spindle and a lamp-and-scale arrangement.

The relationship between the power delivered to the cavity and the amplitude of the mechanical oscillations can be found theoretically, and with knowledge of the specific couple of the quartz fibre suspension and of the damping factor of the mechanical system, measurements of power can be made.

The instrument measures microwave power with an error estimated as not exceeding $\pm 2\%$. The measurement is absolute, since the calibration depends only on measurements of mass, length and time.

LIST OF PRINCIPAL SYMBOLS

B = Capacitive shunt susceptance for detuning element.
E = Electric field strength.
f_0 = Frequency of electromagnetic waves.
Δf = Frequency bandwidth.
Q_0, Q_L = Unloaded and loaded Q-factors of cavity.
J = Moment of inertia.
k = Specific couple of quartz fibre suspension.
M = Mechanical magnification factor.
P = Power.
T = Torque.
W = Electromagnetic energy.
x = Co-ordinate specifying deviation from resonance on universal resonance curve.
θ = Angular displacement.
θ_0 = Angular amplitude of mechanical oscillation.
θ_R = Value of θ at which resonance occurs.
$\Delta\theta$ = Angular bandwidth (corresponding to Δf).

(1) INTRODUCTION

The use of radiation pressure and allied phenomena for the absolute measurement of microwave power has been described previously. None of the instruments hitherto developed, how-ever, is capable of measuring low power levels of the order of a few milliwatts or tens of milliwatts with sufficient accuracy to justify its use as an absolute standard, and it is the purpose of the paper to describe such an instrument suitable for microwaves of about 3 cm wavelength.

This instrument makes use of the principle of mechanical resonance used in the vibrational galvanometer and will be

called the 'vibration wattmeter'. It will be shown that the deflection obtained in an instrument of this type is proportional to $\sqrt{(\text{power/frequency})}$, whereas, for example, the torque-operated wattmeter described by Cullen and Stephenson[1] gives a deflection proportional to (power/frequency). Thus the new instrument has an advantage for measuring small powers at higher frequencies.

On the other hand, unlike the torque-operated wattmeter, which is a reflectionless transmission device, the vibration wattmeter terminates the waveguide with an impedance which varies periodically, and special precautions must be taken to ensure that no error arises from this fact.

(2) PRINCIPLE OF OPERATION

The present instrument employs a rectangular cavity supporting the H_{011} mode as shown in Fig. 1. The cavity is con-
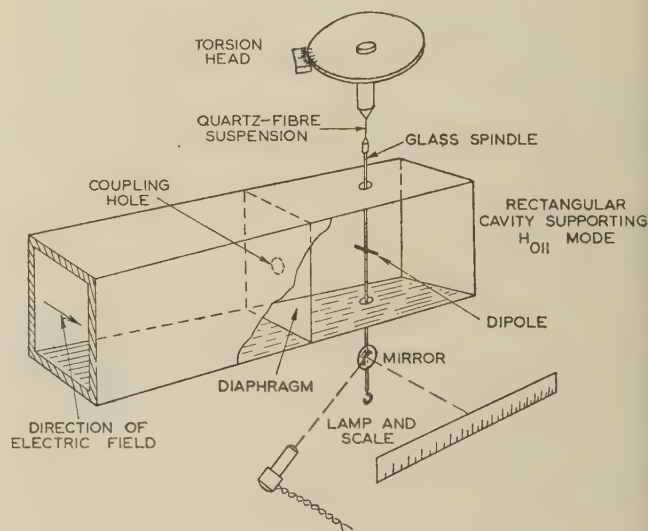
Fig. 1.—General arrangement of simple form of instrument.

structed from standard 0.9 in $\times 0.4$ in waveguide (internal dimensions) and has its broad face in the vertical plane. The glass spindle carrying the dipole passes vertically through the narrow faces of the guide and hangs freely under gravity. The dipole is mounted in a plane at right angles to the axis of the spindle and is located in the centre of the cavity, where the electric field intensity is a maximum. A copper diaphragm with a coupling hole is fixed to the input flange, so forming the closing wall of the cavity. The mechanical force experienced by the rod and the effect of the rod on the resonant frequency of the cavity are now considered separately.

In the absence of the rod the electric field is wholly transverse. If the rod, which acts as a non-resonant dipole, is now introduced at an angle θ [see Fig. 2(a)], there will be a perturbation of the electric field; lines of force will terminate on one end of the

Fig. 2.—Cavity with vibrating dipole.

(a) General arrangement.
(b) Equivalent circuit with lumped constants.

dipole, and similarly lines of force will leave the other end to terminate on the waveguide wall. The dipole therefore acquires an induced dipole moment, and a torque is exerted tending to align it in the direction of the unperturbed electric field, i.e. increasing θ.

In considering the effect of the rod on the resonant frequency of the cavity, it is first noted that the dipole behaves as a variable capacitive shunt susceptance which is a function of the angular position of the dipole [see Fig. 2(b)]. Thus, the resonant frequency of the cavity will depend on the angular position of the dipole. The instrument is adjusted so that the cavity is resonant at the frequency of the energizing power when the dipole is at 45° to the transverse plane. This is also the relaxed position of the quartz fibre. The size of the coupling hole is chosen so that with the dipole at the resonance angle the instrument presents a matched termination to the waveguide.

If now the mechanical system is made to oscillate as a torsional pendulum with $0 < \theta < 90°$, the resonant frequency of the cavity will vary periodically over a range which is dependent upon the amplitude of the mechanical oscillations and upon the characteristics of the dipole. If under these conditions of torsional oscillation the microwave source is switched on, then as the dipole passes through the $\theta = 45°$ position the electromagnetic resonance rapidly builds up and decays, subjecting the mechanical system to an impulsive torque. This impulsive torque acts twice in each cycle, once in an accelerating and once in a retarding sense. If the microwave power is suppressed during the half-cycle in which the retarding impulse would occur, torsional oscillations can be maintained. If the switching is arranged to take place near the extreme positions of the oscillating dipole, when the cavity is almost completely detuned, the precise instant of switching is not important.

The amplitude of these oscillations is clearly dependent on the electric field strength and so is a measure of the microwave power. This is the principle of operation of the instrument.

(3) THEORY OF OPERATION

In this Section the basic equation relating power to the amplitude of the mechanical oscillations of the suspension will be developed.

As in an earlier paper on the use of mechanical forces for microwave power measurement,[3] use is made of the adiabatic invariance of action, proved for an electromagnetic resonator

Fig. 3.—Loss-free resonator and detuning rod.

by McLean.[4] Referring to Fig. 3, consider a loss-free resonator containing a rod (also assumed to be loss-free) which can be rotated about an axis perpendicular to its length, its position being measured by the angle θ shown in the Figure. Suppose that the resonator contains electromagnetic energy W, the field components oscillating at a frequency f_0. Now suppose that the rod is rotated through a small angle $\delta\theta$. Let the change in the frequency of oscillation be δf_0. Then, by the action theorem, there must be a change δW in the energy stored in the electromagnetic field given by

$$\frac{\delta W}{W} = \frac{\delta f_0}{f_0} \quad \ldots \ldots \ldots (1)$$

However, conservation of energy shows that there must be a torque T acting on the rod which can be expressed as

$$\delta W = T\delta\theta \quad \ldots \ldots \ldots (2)$$

By combining eqns. (1) and (2), in the limit

$$T = \frac{W}{f_0}\frac{df_0}{d\theta} \quad \ldots \ldots \ldots (3)$$

In a practical cavity, losses will be present, but provided that the field distribution in the cavity is not significantly affected by these losses, the relationship between total stored energy and torque will not be affected and eqn. (3) is still valid. However, to maintain a constant stored energy W, power must be supplied continuously to the cavity to make up the losses. This can be achieved by the arrangement of Fig. 4. The power required is
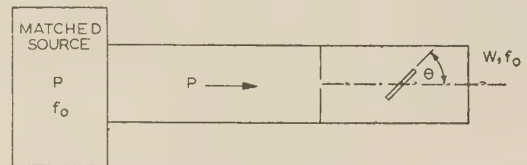
Fig. 4.—Resonator with loss.

conveniently related to the energy stored through the unloaded Q-factor of the resonator:

$$Q_0 = 2\pi \times \frac{\text{energy stored}}{\text{energy lost per cycle}}$$

$$= 2\pi f_0 \frac{W}{P} \quad \ldots \ldots \ldots \ldots (4)$$

By combining eqns. (3) and (4), the torque can be expressed in terms of the power P dissipated in the walls of the cavity and in the dipole itself as

$$T_0 = \frac{PQ_0}{2\pi f_0^2}\frac{df_0}{d\theta} \quad \ldots \ldots (5)$$

This is the torque at resonance, and is the basic equation for the cavity wattmeter described by Bailey.[2] The theory can now be extended to cover the case in which the dipole is not stationary at the peak of the resonance curve, but oscillates in position in the manner described qualitatively in an earlier Section.

16
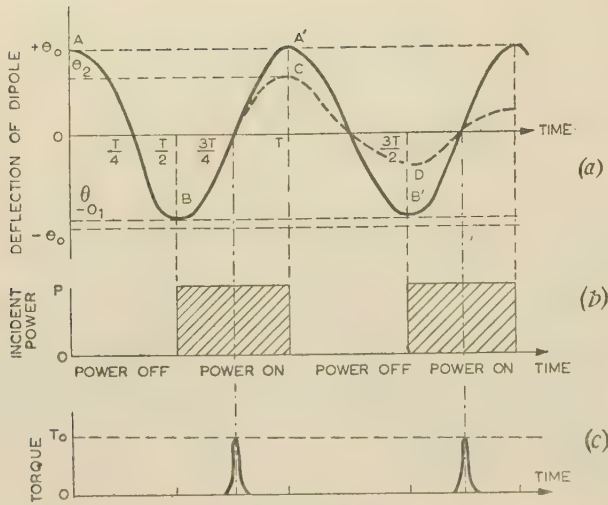
Fig. 5.—Cycle of operation of vibration wattmeter.
(a) Displacement of dipole.
(b) Switching of microwave power.
(c) Torque exerted on dipole.

Fig. 5(a) shows graphically the oscillatory motion of the dipole. In the absence of any microwave power, the dipole will execute damped oscillations as shown by the partly broken curve ABCD. When microwave power is incident on the cavity during the 'forward' swings of the dipole, an impulse will be received by the dipole as it swings through the resonance angle, which is now chosen for convenience as the reference point from which θ is measured. In the steady state, the amplitude of oscillation will be such that the energy lost owing to air resistance and any other causes in one cycle will be just equal to the energy supplied by the impulse received during that cycle.

These ideas may now be applied to calculate the amplitude of oscillation in terms of the microwave power. As the dipole swings, it varies the resonant frequency of the cavity. If the cavity is matched to the guide at resonance and the source is matched to the guide, the electric field E in the cavity is given in terms of its value E_0 at resonance by the equation

$$E = \frac{E_0}{1 - j2Q_L\frac{\delta f_0}{f_0}} \quad \ldots \ldots \quad (6)$$

In this formula, δf_0 is the increment in resonant frequency and Q_L is the loaded Q-factor of the cavity. Since the source is matched to the guide and the cavity is also matched to the guide, the loaded Q-factor is just half the unloaded Q-factor, i.e.

$$Q_L = \tfrac{1}{2}Q_0 \quad \ldots \ldots \quad (7)$$

Since the torque acting on the dipole is proportional to the square of the magnitude of the electric field, the torque T can be expressed in terms of its value T_0 at resonance as

$$T = \frac{T_0}{1 + \left(Q_0\frac{\delta f_0}{f_0}\right)^2} \quad \ldots \ldots \quad (8)$$

The torque is significant only near resonance, and an adequate approximation for δf_0 in terms of the angle θ measured from the resonance angle is

$$\delta f_0 = \frac{df_0}{d\theta}\theta \quad \ldots \ldots \quad (9)$$

Combining eqns. (8) and (9),

$$T = \frac{T_0}{1 + \left(\frac{Q_0}{f_0}\frac{df_0}{d\theta}\theta\right)^2}$$

or, using eqn. (5),

$$T = \frac{P}{2\pi f_0}\frac{\left(\frac{Q_0}{f_0}\frac{df_0}{d\theta}\right)}{1 + \left(\frac{Q_0}{f_0}\frac{df_0}{d\theta}\theta\right)^2} \quad \ldots \ldots \quad (10)$$

The work done on the dipole is given by

$$\delta W = \int_{-\infty}^{+\infty} T d\theta \quad \ldots \ldots \quad (11)$$

The limits should strictly be the angular limits of the mechanical oscillations, but when these limits are large in comparison with the angular bandwidth of the cavity, the integral may be taken from $-\infty$ to $+\infty$ without significant error. The angular bandwidth is defined as the angle through which the dipole rotates to alter the resonant frequency of the cavity by an amount equal to the frequency bandwidth Δf. Putting

$$x = \frac{Q_0}{f_0}\frac{df_0}{d\theta}\theta$$

and combining eqns. (10) and (11),

$$\delta W = \frac{P}{2\pi f_0}\int_{-\infty}^{+\infty}\frac{dx}{1 + x^2}$$

or

$$\delta W = \frac{P}{2f_0} \quad \ldots \ldots \quad (12)$$

This is the energy gained per cycle.

Now consider the energy loss due to damping. Natural oscillations of the system will decay exponentially thus:

$$\theta = \theta_0\varepsilon^{-\alpha t}\cos\Omega t \quad \ldots \ldots \quad (13)$$

and in particular, the amplitude after one half-cycle, θ_1, is given by

$$\theta_1 = \theta_0\varepsilon^{-\alpha\tau/2} \quad \ldots \ldots \quad (14)$$

After a complete cycle, the amplitude θ_2 is

$$\theta_2 = \theta_0\varepsilon^{-\alpha\tau} \quad \ldots \ldots \quad (15)$$

At the beginning and end of the first half-cycle the energy is potential only, and so the energy lost in this half-cycle is

$$\tfrac{1}{2}k\theta_0^2 - \tfrac{1}{2}k\theta_1^2 \quad \ldots \ldots \quad (16)$$

Free oscillations have been considered, but under conditions of maintained oscillations alternate half-cycles, e.g. AB, A'B', ... in Fig. 5(a) are essentially free oscillations, and the loss of energy due to air resistance can be calculated from relation (16) for these half-cycles.

For the other half-cycles, BA', etc., the loss of energy due to air resistance must be very nearly the same. It will be shown later that the error in assuming that it is exactly the same can be made negligible (see Section 13).

By making this assumption, the energy $\delta W'$ lost owing to air resistance is given by

$$\delta W' = k(\theta_0^2 - \theta_1^2)$$

or, using eqn. (14),

$$\delta W' = k\theta_0^2(1 - \varepsilon^{-\alpha\tau}) \quad \ldots \ldots \quad (17)$$

In the steady state of maintained oscillations the energy lost in each cycle in friction, given by eqn. (17), is equal to the energy gained in each cycle from the electromagnetic field, given by eqn. (12). Hence, the steady-state amplitude θ_0 is given by

$$\theta_0^2 = \frac{P}{2f_0 k} \frac{1}{1 - \varepsilon^{-\alpha\tau}} \quad . \quad . \quad . \quad . \quad (18)$$

In practice, the factor $\varepsilon^{-\alpha\tau}$ is determined by noting the amplitude of successive swings in natural oscillations of the system. From eqn. (15)

$$1 - \varepsilon^{-\alpha\tau} = \frac{\theta_0 - \theta_2}{\theta_0}$$

The quantity on the right-hand side of this equation is the fractional decrease in amplitude per cycle.

If we define a quantity M by the equation

$$M = \frac{\theta_0}{\theta_0 - \theta_2}$$

so that M can be thought of as the 'magnification' of the mechanical resonance, eqn. (18) becomes

$$\theta_0^2 = \frac{P}{2f_0} \frac{M}{k} \quad . \quad . \quad . \quad . \quad . \quad (19)$$

This is the basic equation of the system, and is sufficiently accurate throughout the range of powers for which the instrument is useful.

It can be seen from eqn. (12) that the mechanical energy derived from the electromagnetic field is proportional to the area under the tuning curve, and accurate measurements of power depend on the resonance curve having its theoretical shape. If the Q-factor of the cavity were increased, the instantaneous electric field strength would be correspondingly greater, but the resultant reduction in bandwidth would keep the impulses delivered to the suspension at a constant magnitude. Similarly, if the dipole were lengthened, the instantaneous torque would be increased, but this again would be compensated by an increased rate of detuning. Thus, for a given power, the amplitude of the mechanical oscillations is independent of both the Q-factor of the cavity and the characteristics of the dipole, as can be seen in eqn. (19).

The instrument might be regarded as a microwave reciprocating engine, converting some of the microwave power into mechanical power. During the short period when the cavity passes through resonance most of the available power from the source is absorbed in the walls of the cavity. Only a few microwatts of electrical power are used to drive the pendulum.

(4) ELECTRICAL AND MECHANICAL DESIGN

From the theory it can be seen that the two most important considerations which must be fulfilled for the accurate operation of the instrument are as follows:

(a) The detuning of the cavity must be linear for a sufficiently large range to prevent distortion of the resonance curve.

(b) The angular bandwidth of the resonance curve must be small compared with the amplitude of the mechanical oscillations. This ensures that the energy is imparted to the mechanical system impulsively and is one of the factors governing the lowest power that can be measured.

Experiments were conducted to investigate the characteristics of a rotating capacitive element in a waveguide, and the shunt susceptance was found to obey, to a very close approximation, the law

$$B = B_{max} \sin^2 \theta$$
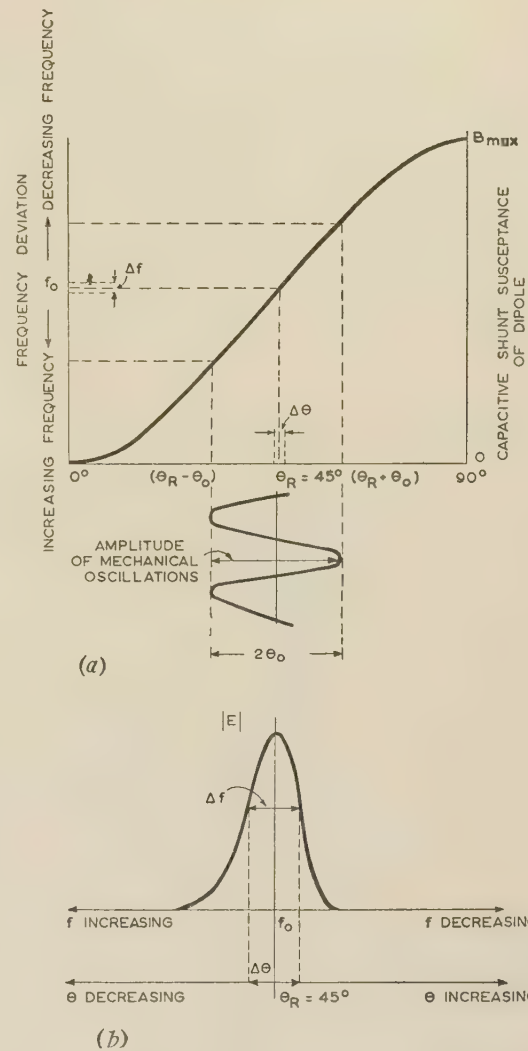
Fig. 6.—Tuning characteristics of cavity with rotating dipole.
(a) Resonance frequency (f)-rotation (θ) characteristic.
(b) Tuning curve.

where B = Capacitive susceptance of the element.
θ = Angular position of the element as in Fig. 2(a).

This law has also been predicted for small elements by theoretical investigation. Thus there is a linear relationship between the frequency shift and the rotation of the dipole over a considerable range centred on the $\theta = 45°$ position [see Fig. 6(a)]. It is also to be noted that the maximum torque occurs at $\theta = 45°$.

For the angular bandwidth [see Fig. 6(b)] to be small compared with the amplitude of the oscillation it is necessary to have a fairly high Q-factor for the cavity and a large change of frequency per degree rotation of the dipole. It is obvious from Fig. 6(a) that the larger the dipole the greater the slope $df/d\theta$. The largest dipole that can be suitably accommodated in the cavity is about 10 mm long. While this provides the maximum possible rate of detuning, it can also introduce some undesirable effects. From a study of eccentric vanes in waveguides it was found that the capacitive shunt susceptance of the vane increases very rapidly as the vane approaches the wall of the guide. Thus any effects which cause the dipole to move in a manner which is not purely torsional are likely to affect the linearity of detuning.

These effects can arise from external vibrations or from mechanical asymmetry of the moving parts. Electrical asymmetry of the dipole may also give rise to translation as well as the desired purely torsional oscillations, especially if the frequency of torsional oscillation is a harmonic of the pendulum frequency. The larger the dipole the more serious these problems become. A dipole length of about 8 mm is a satisfactory compromise.

Thus the adjustment of the angular bandwidth is most suitably effected by controlling the Q-factor of the cavity. While a very high Q-factor produces a desirably small bandwidth, it also makes the problems associated with the frequency instability of the source more acute, as will be discussed in the next Section.

The prototype instrument is shown in more detail in Fig. 7.
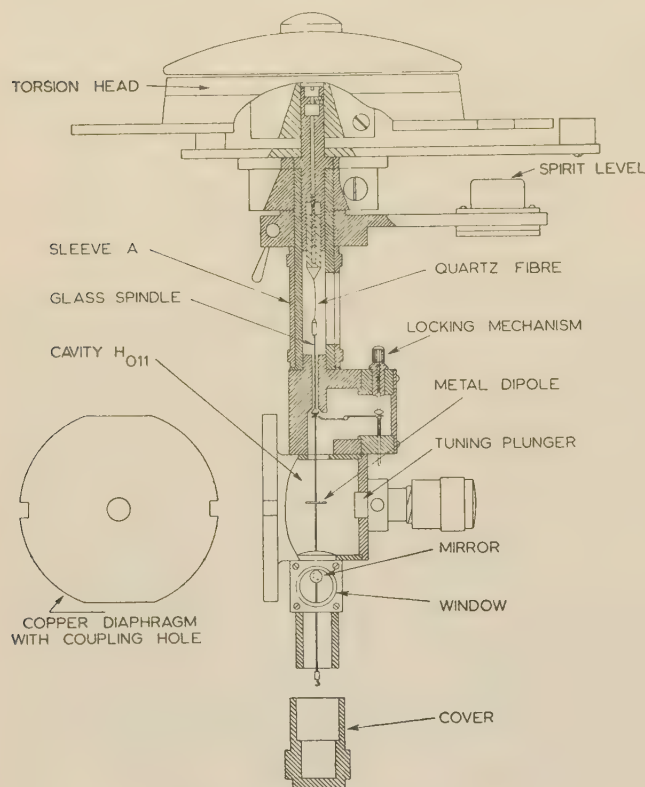
Fig. 7.—The vibration wattmeter.

The glass spindle, having a diameter of 0·02 in, passes through the waveguide walls and carries the dipole. The hole in the lower wall is large enough to allow the dipole to pass, and when the suspension is in position a small plug is inserted to prevent loss of electromagnetic energy which would otherwise be radiated into the mirror housing. The upper end of the spindle carries a small Dural cap suitably shaped for the fixing of the 10-micron quartz fibre, and it also carries the cone which seats into a conical recess when the spindle is lifted. This clamps the suspension and relieves the tension in the quartz fibre. The clamping mechanism is housed immediately above the cavity; Fig. 7 shows it in the locked position. To the lower end of the suspension is fitted a mirror of 100 cm focal length, and at the extremity is fitted a Dural hook which is used for the attachment of rods in the mechanical calibration experiments. A cover is fitted at the bottom of the instrument to exclude draughts while power measurements are made, and this can also be used as an oil dash-pot to steady the suspension while the electrical characteristics are being measured.

The spindle with its attachments is assembled in a jig and is inserted through the lower end of the instrument. The plug is then fitted and the clamping mechanism locked; this holds the suspension steady while the quartz fibre is attached. Access for fixing the fibre is made through a slot in the rotating sleeve A.

The vertical position of the dipole in the cavity can be adjusted by raising or lowering the top spigot, to which the fibre is fixed, by a mechanism which is integral with the torsion-head spindle.

(5) CALIBRATION

The calibration of the instrument for absolute measurement of power involves the determination of two mechanical quantities; first, the constant of proportionality between the torque and deflection of the quartz fibre suspension and secondly the damping factor of the moving parts. Although no electrical constant has to be determined for the evaluation of power, it is necessary to check the electrical performance to ensure that the operation of the instrument conforms to the assumptions made in the theory.

(5.1) Mechanical Calibrations

The specific couple exerted by the fibre was found by fixing various rods of known moment of inertia to the suspension, and measuring the periodic time of the system for small oscillations. Corrections were made for the effect of air damping.

The specific couple was calculated from the expression

$$k = \frac{J}{t_1^2 - t_0^2}\left[4\pi^2 + \left(\log_\varepsilon \frac{A_n}{A_{n+1}}\right)^2\right]$$

where J = Moment of inertia of the rod.
  t_0 = Periodic time of the suspension alone.
  t_1 = Periodic time of the suspension with the rod attached.
A_n, A_{n+1} = Amplitudes of successive swings.

The mechanical damping (the inverse of the magnification factor M) is defined as $(\theta_0 - \theta_2)/\theta_0$ and was determined by recording the amplitudes of successive swings of the free oscillations of the suspension. About a centimetre length of 10-micron quartz fibre is used for the suspension, and this has a torsional stiffness of about 3×10^{-2} dyne-cm per radian. The mechanical magnification factor M is of the order of 20.

It may be assumed that the constant M is almost entirely due to air friction and is affected to a negligible extent by internal hysteresis losses in the quartz fibre. For a given oscillating element, M will be inversely proportional to the air viscosity, and it must be considered how this will change with temperature and pressure. Maxwell showed in 1860 that on the kinetic theory the viscosity of a gas does not depend on its pressure. This is in accordance with experiment, provided that the mean free path of the gas molecules is small in comparison with the dimensions of the apparatus, and may be assumed to be true here. The variation with temperature is given by a formula due to Sutherland, accurate at temperatures above the critical value and at pressures such that Boyle's law is obeyed. The formula is

$$\eta_t = \eta_{273}\left(\frac{273 + C}{t + C}\right)\left(\frac{t}{273}\right)^{3/2}$$

For air the constant C has the value 117. Calculation shows that if the instrument is calibrated at 20° C the apparent power at any temperature t° C should be multiplied by the factor

$$1 + 0\cdot00268(t - 20)$$

to find the true power.

Thus a change in temperature from 20° C to 15° C requires a correction of 1·3% to the apparent power to find the true power.

The effect of the water-vapour content of the air on the frictional damping has been investigated. Experiments were carried out on a suspension mounted in a closed tube, where the amount of water vapour could be controlled, and no significant effects were detected.

However, the comparison tests were carried out at the same room temperature at which the substandard thermistor bridge had been calibrated, and the value of M was checked several times during the comparisons.

(5.2) Electrical Characteristics

Experiments were carried out to check the following points:

(*a*) The resonance angle of the dipole is 45° at the frequency at which power is to be measured, and that at this setting the instrument presents a matched termination to the waveguide run.

(*b*) There is only one resonance within the angular range to be swept by the dipole.

(*c*) The detuning curve of resonant frequency of the cavity against angular displacement of the dipole is linear over a range of not less than ten times the bandwidth.

(*d*) The angular bandwidth $\Delta\theta$ is small compared with the smallest value of θ_0 to be measured (see Section 13), where

$$\Delta\theta = \frac{f_0}{Q_L}\frac{d\theta}{df}$$

With careful designing of the instrument and with a cavity working in the fundamental mode these conditions are not difficult to satisfy.

(6) FREQUENCY STABILITY OF THE SOURCE

The inherent accuracy of the vibration wattmeter may now be discussed. However, the accuracy realizable in practice depends on the frequency stability of the source, and this must be considered before a realistic assessment of the accuracy of power measurement can be made.

The theoretical relationship between the amplitude of mechanical oscillations and the microwave power flow is only valid if the frequency of the source remains constant. For the greater part of the cycle of operation of the instrument the cavity is far from its resonant condition, and it presents a termination which is almost a short-circuit in the plane of the diaphragm. As the cavity passes through resonance, frequency and amplitude pulling will occur if the source is closely coupled to the cavity. The magnitude of this frequency shift is a major factor determining the ultimate accuracy of the measurements.

From the theoretical analysis it is evident that the energy imparted to the dipole is proportional to the area under the resonance curve. It is therefore imperative that the tuning curve should correspond faithfully to the shape of the theoretical curve. The ordinate of the resonance curve represents the magnitude of the electric field strength within the cavity, and the square of this is therefore a measure of the instantaneous torque experienced by the dipole. Thus, as the dipole is urged forward through a small angle $\delta\theta$ owing to the influence of the electric field, the work done on the dipole is $T\delta\theta$ (see Fig. 8). Hence the total work done per mechanical cycle is proportional to the area $\int_{-\infty}^{+\infty} Td\theta$ under the tuning curve. The Figure illustrates a typical oscillogram which has been obtained when amplitude and frequency pulling have occurred. The area under the theoretical curve (solid line) has been increased by the amount shown shaded, with an apparent increase in measured power corresponding to the excess area. It becomes evident, therefore,
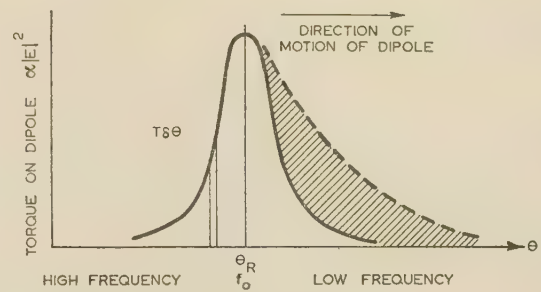
Fig. 8.—Effect on resonance curve of frequency and amplitude pulling.

that any frequency pulling that occurs must be small compared with the bandwidth.

The bandwidth of the resonance curve is determined by the Q-factor of the cavity and the detuning characteristics of the dipole. As neither of these quantities affects the measurements of power, within the limits described in Section 3, it is advantageous to keep the Q-factor and the detuning rate of the dipole fairly low to ensure a reasonably large bandwidth and so make the frequency stabilization problem less critical.

This adjustment is a compromise, since it is a necessary condition for the correct measurement of power that the angular bandwidth $\Delta\theta$ should be small compared with the semi-amplitude θ_0. It is shown in eqn. (22) that if the angular amplitude of swing θ_0 is plotted against the square root of the microwave power, the result will be a straight line through the origin if the angular bandwidth is zero. For a small but finite angular bandwidth $\Delta\theta$ the result will be a straight line with a negative intercept $\Delta\theta/2\pi$. The simple proportionality between θ_0^2 and power no longer holds exactly, but for a given accuracy it holds adequately provided that the microwave power exceeds a certain minimum value. It is this condition which determines the ultimate low-power limit.

A quantitative analysis of these effects is complicated, but the solution of an idealized case quickly yields the limits within which the frequency must remain constant for a given maximum error in the measured power flow.

For a given frequency-shift the maximum possible increase in area under the resonance curve would occur if the frequency of the source shifted at exactly the same rate at which the dipole detuned the cavity, and if this occurred at the maximum height of the tuning curve as indicated in Fig. 9.
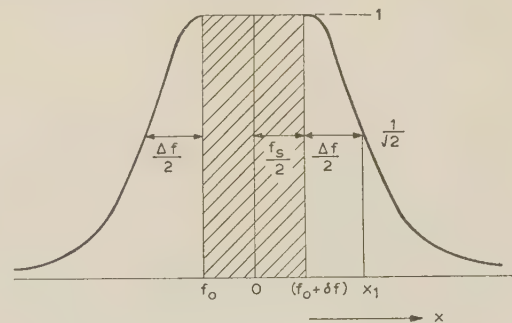
Fig. 9.—An idealized case of frequency pulling.

The total area under the undistorted normalized resonance curve is

$$2\int_0^\infty \frac{dx}{1+x^2} = \pi$$

Now suppose the area must not increase by an amount greater

than $N\%$, then the smallest frequency shift which would produce this additional area is f_s, where

$$f_s = \frac{\Delta f}{0 \cdot 643\,4} \frac{N\pi}{200}$$

A typical value for the loaded Q-factor of the instrument is 3 000, and the frequency bandwidth at a free-space wavelength of $3 \cdot 2$ cm is $3 \cdot 12$ Mc/s. This corresponds to an angular bandwidth of approximately one-quarter of a degree as the slope of the detuning curve is $8 \cdot 2$ Mc/s per degree.

Thus a total frequency shift f_s of 76 kc/s from the operating frequency f_0 could cause an error of 1% in the measured power. In practice, the frequency pulling does not occur at the peak of the resonance curve, but is distributed as shown in Fig. 8, and a frequency shift much greater than 76 kc/s would be required to contribute the calculated error.

To fulfil the above conditions of frequency stability it was necessary to provide good isolation between the instrument and the source. It was found that an attenuating pad of greater than 25 dB power loss was essential to limit the frequency and amplitude pulling to an acceptable amount if powers of the order of a few milliwatts were to be measured. By using either an attenuator or a directional coupler to provide the necessary isolation, an input power of several watts would be required. This makes impossible a direct measurement of low powers from a source such as a klystron, although the absolute calibration of a thermistor can be made.

This limitation was overcome by using a non-reciprocal attenuator employing ferrite. The attenuator was of the resonance absorption type using a transverse steady magnetic field, the wave being propagated in a rectangular guide supporting the H_{01} mode.

By careful adjustment at a spot frequency the attenuation to the reflected wave was about 27 dB, while the attenuation in the forward direction was less than $1 \cdot 5$ dB. A variable resistive-card attenuator was used in conjunction with the non-reciprocal attenuator for the adjustment of power level, and adequate stability of the source was obtained.

The frequency was monitored by means of a cavity wavemeter which was accurate to an absolute value of 1 part in 10^4, and by tuning the wavemeter to the region of greatest slope of its response curve, changes in frequency of the order of 40 kc/s could be detected easily. Before making measurements with the instrument, the waveguide was terminated with a short-circuiting plunger which was moved through a guide half-wavelength to ensure that no frequency or amplitude pulling could be detected.

(7) ACCURACY

The calibration procedure is, of course, subject to experimental errors, both systematic and random. The experimental values for the specific couple of the quartz fibre, obtained by using five rods of different lengths and masses, show a standard deviation of $0 \cdot 18\%$, and the determination of the mechanical magnification factor M gives a standard deviation of $0 \cdot 30\%$.

The combination of these two random errors gives an overall standard deviation of $0 \cdot 35\%$. Taking three times the standard deviation as a reasonable error limit, it is unlikely that the random errors in calibration could exceed $\pm 1 \cdot 05\%$.

Since the Q-factor of the cavity and the characteristics of the dipole need not be known, beyond the knowledge that they are of suitable magnitudes, no random or systematic errors are associated with these electrical measurements.

Systematic errors in the mechanical calibration may arise from the following causes:

(a) Incorrect graduation of scales used for measuring the length of the rods.
(b) Inaccuracies in weights used in weighing the rods.
(c) Errors in timing the period of oscillation.
(d) Incorrect graduation of the scale used to determine M.

No direct evaluation of errors is possible in cases (a), (b) and (d). However, good agreement has been achieved by checking with apparatus and techniques employed previously,[1] and the internal consistency of the measurements gives reasonable assurance that these errors are probably negligible.

The stop watch used for timing was checked over a long period against the Post Office speaking clock TIM. Errors in calibration arising from this source were certainly less than $0 \cdot 02\%$.

In using the instrument for measurements of power, further systematic errors may arise:

(e) Imperfect matching.
(f) Errors in measurement of frequency.
(g) Errors caused by frequency pulling.
(h) Incorrect graduation of scale used for measurement of deflection, and errors in measuring mirror-to-scale distance.

When the instrument was adjusted to the resonant condition, it presented a termination whose voltage standing-wave ratio was better than $0 \cdot 995$. The measurement of this standing-wave ratio presented some difficulty because of the movement of the dipole under the influence of the power supplied by the oscillator of the waveguide test bench. For this reason the measurement was made with the dipole damped at $45°$, great care being taken to ensure that the dipole was in the correct position. The generator was also matched.

The error introduced by these mismatches depends on the length of guide between the source and the cavity. The effect of mismatch is twofold: the power delivered to the cavity is modified (this is a very small effect), and, more important, the relationship between the loaded and unloaded Q-factors is modified. If the standing-wave ratios of the cavity and source are S_1 and S_2, then in the worst case the formula for the square of the deflection [eqn. (19)] must be multiplied by the factor $2/(1 + S_1 S_2)$. In our case, S_1 and S_2 both exceed $0 \cdot 995$, so that the multiplying factor cannot be greater than $1 \cdot 005$, i.e. the maximum error due to this cause is $\frac{1}{2}\%$.

The wavemeter used to determine the frequency was accurate to an absolute value of 1 part in 10^4 and would introduce an error of $0 \cdot 01\%$. It was estimated that the error introduced by frequency pulling would not exceed $0 \cdot 3\%$.

The systematic error in the scale arrangement for measuring the deflection of the suspension is estimated as not exceeding $0 \cdot 1\%$.

Experiments have shown that the levelling of the instrument is not critical and does not introduce an error into the calibrations.

In the statistical combination of the standard deviations with the estimated error limits, the estimated errors are taken as equivalent to three times the standard deviation (e.g. in the estimated mirror-to-scale distance where a standard deviation is not available).

The total inherent error of the instrument is unlikely to exceed $1 \cdot 6\%$, and in addition the error due to frequency pulling will not exceed $0 \cdot 30\%$.

Finally, it may be said that the results show that errors due to all causes in the measurement of power with this instrument are not likely to exceed $\pm 2\%$.

(8) EXPERIMENTAL PROCEDURE

In order to maintain mechanical oscillations the electromagnetic energy is suppressed, on alternate mechanical half-cycles, by an automatic switching circuit which is triggered by a light spot passing a photocell. The recent development of ferrite

devices suggests the possibility of using a ferrite switch which would act as the modulator and isolator combined. This scheme has the advantage that the microwave source would feed into a matched load throughout the switching cycle.

To measure power, the movement is set in motion, and the switching circuit automatically adjusts itself to the correct phasing. The oscillations then build up to a steady magnitude, and the amplitude is read from a curved scale set 100 cm from the mirror. The most satisfactory method of reading the amplitude of oscillations was found to be by the use of marker cards which could be adjusted until their hair lines coincided with the limits reached by the moving hair line projected with the light spot. The amplitude could then be measured on the scale.

It was found helpful to monitor the instrument while it was in operation. This was done by displaying on a cathode-ray oscilloscope the resonance curve derived from a standing-wave indicator probe which was located immediately before the instrument at a minimum in the standing wave when the cavity was far from resonance. On the second beam were displayed the slow-speed square-wave modulation of the power source, and marker pips derived from the automatic timing circuits. With this display it is possible to make final adjustments while the instrument is working, and it also makes easier the detection of faulty operation.

The centre about which the suspension oscillates is displaced from the static centre by an amount proportional to the mechanical magnification factor M, where M is defined as $\theta_0/\delta\theta_0$.
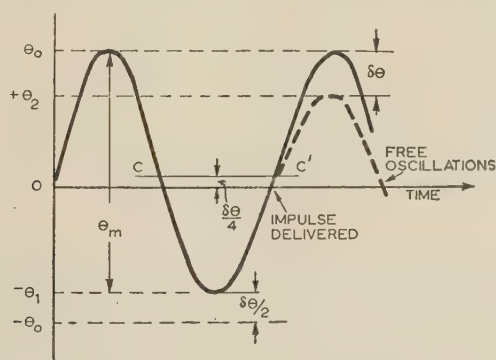
Fig. 10.—The effect of damping on the measured amplitude θ_m.

It can be seen from Fig. 10 that the new centre is about the line CC^1 and is displaced by an amount $\frac{1}{4}\delta\theta$.

The instrument can readily be realigned to compensate for this asymmetry with the aid of the monitoring oscilloscope. However, a correction term must be used to determine the true semi-amplitude θ_0. The apparent amplitude θ_m as calculated from the scale readings is given by

$$2\theta_m = 2\theta_0 - \tfrac{1}{2}\delta\theta$$

and

$$\frac{\theta_0}{\delta\theta_0} = M$$

Hence, the true semi-amplitude is

$$\theta_0 = \theta_m\left(\frac{4M}{4M-1}\right) \simeq \theta_m\left(1 + \frac{1}{4M}\right)$$

Thus if M is 20, the correction is $1\cdot25\%$.

(9) COMPARISON WITH THERMISTOR BRIDGE

Comparison tests have been carried out at a wavelength of $3\cdot18$ cm, in which the output power of a continuous-wave

klystron has been measured by a thermistor bridge and the vibration wattmeter. The thermistor bridge had been previously calibrated against a water calorimeter and the torque vane wattmeter,[5] and the mean thermistor bridge power was estimated to be accurate to about $\pm2\%$. The comparisons were made over a range 7–29 mW using a calibrated glass-vane attenuator in conjunction with the thermistor bridge, which was operated at the milliwatt level. Because of the characteristics of the vibration wattmeter, a substitution method was employed to compare the two instruments, and all practical precautions were taken to ensure that the delivery of power was constant during each test. The results are given in Table 1, in which

P_w = Power indicated by the vibration wattmeter.
P_T = Power indicated by the thermistor bridge.

Table 1

COMPARISON BETWEEN VIBRATION WATTMETER AND THERMISTOR BRIDGE

P_w	P_T	$\dfrac{(P_w - P_T)}{P_T} \times 100$
mW	mW	%
28·9	27·8	+4·3
27·2	26·7	+1·9
26·8	25·4	+5·5
26·5	26·1	+1·5
23·4	24·7	−5·3
22·2	21·8	+1·8
21·5	21·7	−0·9
21·0	21·0	0·0
19·2	19·2	0·0
18·4	18·5	−0·5
15·9	15·5	+2·6
14·2	14·4	−1·4
14·0	14·3	−2·1
12·9	13·0	−0·8
11·3	11·3	0·0
8·8	8·8	0·0
8·4	8·9	−5·6
7·8	8·2	−4·9
7·7	7·8	−1·3
7·1	7·2	−1·4

The average of the percentage difference is $-0\cdot3\%$, which shows that there is no significant systematic discrepancy, and the results are mostly within the estimated limits of error for either instrument.

Random errors in the comparison are introduced mainly by the effect of external vibrations. Fourteen of the twenty readings lie within a range of $\pm2\frac{1}{2}\%$, and the rather large discrepancies shown by the remaining six readings are known to be due to excessive vibration, which affects the delicate suspension of the vibration wattmeter.

The comparison tests were carried out in daytime, when interference due to external vibrations is greatest. Other tests, employing a calibrated crystal and attenuator as a substandard, have been carried out in the early hours of the morning, when external vibrations are at a minimum, and it has been found that the scatter has not exceeded $\pm1\%$.

The external vibrations experienced were quite severe, owing to the unfavourable location of the laboratory near a busy main road.

(10) CONCLUSIONS

The vibration wattmeter will measure c.w. microwave power from about 5 to 30 mW to an accuracy of about $\pm2\%$ in the 3 cm band. There is no difficulty in extending its power range to the

100 mW level and above, by increasing the stiffness of the quartz fibre. The lower limit for accurate power measurement is set partly by the difficulty of reading the small deflections and partly by external vibrations. It could be made less sensitive to external vibrations by employing a dynamically-balanced double-ended suspension.

The ultimate lower limit is reached when the deflection becomes comparable with the angular bandwidth. The present instrument is tunable over the range $3 \cdot 10 – 3 \cdot 25$ cm, but there is no fundamental difficulty in designing an instrument for use over a wider band.

The instrument with its auxiliary apparatus is intended for laboratory use, but is quite portable.

As pointed out in the Introduction, the deflection of this instrument is proportional to the square root of the power and inversely proportional to the square root of the frequency, and this characteristic is a promising feature for the application of the instrument to measurements at higher frequencies.

At present work is being carried out on an instrument for 8 mm wavelengths.

(11) ACKNOWLEDGMENTS

(12) REFERENCES

(1) CULLEN, A. L., and STEPHENSON, I. M.: 'A Torque-Operated Wattmeter for 3 cm Microwaves', *Proceedings I.E.E.*, Monograph No. 42 M, July, 1952 (**99**, Part IV, p. 294).
(2) BAILEY, R. A.: *ibid.*, Monograph No. 138 R, June, 1955 (**103** C, p. 59).
(3) CULLEN, A. L.: 'A General Method for the Absolute Measurement of Microwave Power', *ibid.*, Monograph No. 23 M, February, 1952 (**99**, Part IV, p. 112).
(4) MACLEAN, W. R.: 'The Resonator Action Theorem', *Quarterly Journal of Applied Mathematics*, 1945, **2**, p. 329.
(5) BAILEY, R. A., FRENCH, H. A., and LANE, J. A.: 'The Comparison and Calibration of Power-Measuring Equipment at Wavelengths of 3 cm and 10 cm', *Proceedings I.E.E.*, Paper No. 1689 R, September, 1954 (**101**, Part III, p. 325).

(13) APPENDIX

(13.1) Assumptions about Energy Loss

In deriving eqn. (18) it was assumed that the loss of energy due to air damping during the forced swing (BA′ in Fig. 5) is the same as during the subsequent free swing (A′B′). The error due to this assumption can be estimated by considering the case of an impulsive force, i.e. infinite Q-factor. In this case the exact solution can be shown to be

$$\theta_0^2 = \frac{P_0}{\omega_0} \frac{2\pi}{k} \frac{\exp\left[-\frac{\alpha}{\Omega}\left(\pi - 2\tan^{-1}\frac{\alpha}{\Omega} \right) \right]}{1 - \exp\left(-\frac{4\pi\alpha}{\Omega} \right)} \quad . \quad . \quad (20)$$

This equation can be rewritten in terms of the mechanical magnification factor M, and the first-order correction factor for eqn. (19) is found to be

$$1 - \left(\frac{1}{8} - \frac{1}{2\pi^2} \right) \frac{1}{M^2} \quad . \quad . \quad . \quad . \quad (21)$$

If $M > 10$, the error in using eqn. (19) is less than $0 \cdot 075\%$. Provided that the Q-factor is sufficiently large, therefore, the assumption about air-damping is probably quite adequate.

(13.2) Assumption about Integration Limits

Another assumption made in deriving eqn. (19) is that the limits can be taken as infinite in evaluating the integral leading to eqn. (12). If the correct limits are retained, it is found that the first-order correction to eqn. (19) leads to

$$\theta_0 = \sqrt{\left(\frac{P_0}{2f} \frac{M}{K} \right)} - \frac{\Delta\theta}{2\pi} \quad . \quad . \quad . \quad (22)$$

When the 'angular bandwidth' $\Delta\theta$ is small, eqn. (19) can be used together with a correction factor

$$1 - \frac{\Delta\theta}{\pi\theta_0} \quad . \quad . \quad . \quad . \quad . \quad (23)$$

In this case the angular bandwidth is $0 \cdot 37°$, so that if $\theta_0 = 10°$ the error is about 1%, and is proportionally less for larger amplitudes of swing. A correction can easily be applied where significant.

(13.3) Independence of Foregoing Assumptions

The further assumption is made that the approximation used for the air-damping losses discussed in Section 13.1, which was justified only for infinite Q-factor, is still valid for large but finite Q-factor. This seems intuitively highly probable to the authors, though no formal proof has been obtained.

The form of eqn. (22) has, however, been verified experimentally, and the right order of magnitude of intercept obtained. Thus it is safe to assume that no significant error arises from assuming that the effects of air-damping and finite Q-factor are largely independent.

DOUBLE-SLAB ARBITRARY-POLARIZATION SURFACE-WAVE STRUCTURE

By ROBERT E. PLUMMER, B.S., and ROBERT C. HANSEN, Ph.D.

SUMMARY

A single earthed dielectric slab cannot propagate both polarizations of surface waves with equal velocities, so a more complex structure such as a double-layer slab is needed. The nature of surface waves on 2-layer earthed-slab structures is studied quantitatively, and conditions are derived that allow for equal velocities of both polarizations; a situation that will allow a wave of arbitrary or circular polarization to propagate. The upper slab must have a greater dielectric constant, and the square of the inverse velocity ratio (of the surface wave compared with free space) is constrained to lie in a range only slightly greater than that between the two dielectric constants. Extensive numerical data for the design of these arbitrary polarization structures are included.

LIST OF SYMBOLS

$k = \omega\sqrt{(\mu_0\epsilon_0)}$ = Free-space wave number.

μ_0, ϵ_0 = Free-space permeability and permittivity, respectively.

ϵ_1, ϵ_2 = Relative permittivities of the first and second layers, respectively.

γ, δ, α = Transverse wave numbers of the first, second and air layers, respectively.

β = Wave number (phase-change coefficient) in the z-direction.

a, b = Thicknesses of the first and second layers, respectively.

(1) INTRODUCTION

There has been much written in the literature of recent years on the subject of surface waves. The question of the applicability of surface waveguides to low-loss microwave transmission systems and to low-silhouette antenna structures has prompted renewed attention to a long-known phenomenon.

The two surface-wave structures that have received the greatest attention are the flat corrugated conducting surface and the plane dielectric-clad conductor. Both of these types of surface waveguide are limited to a single polarization only; the corrugated surface can support only TM waves, and a single earthed dielectric slab cannot propagate both TM and TE waves or polarizations with the same velocity.

For some applications it is desirable that the transmission system or antenna be capable of supporting a wave of arbitrary polarization.* Propagation of a wave of arbitrary polarization over a surface-wave system requires that both TM and TE waves possess the same propagation coefficient. Two practical surface-wave structures are known for which this requirement can be met. These are the single-layer earthed dielectric slab with mode filter and the double-layer earthed dielectric slab. Hansen[1] has investigated the single earthed slab with mode filter, and has shown that the phase velocities of the two polarizations can be independently adjusted with such a system. Two-layer or multi-layer dielectric slabs for the guiding of surface waves have been studied by several groups. Ehrenspeck *et al.*[2] have given the general equations for an *n*-layered slab. Karbowiak[3] discusses the characteristics of the TM wave over an *n*-layered slab in which the layer thicknesses are very thin. Jones and Cohn[4] consider 2-layer earthed slabs, but they are primarily interested in the dominant TM wave. This latter report gives graphical data for the case of an upper layer of polystyrene and a lower slab of various foam materials. These cases are completely explored and curves of external power flow as well as of surface electric field are given.

More abstruse work by Kay and Zucker[5,6] relates the excitation by an arbitrary source over multi-layered media to the far field of the source in free space. The first to propose the double-layer dielectric slab for circular polarization appears to have been Zucker.[7] His paper, which is an excellent summary of the state of the art, reports work in progress at the Air Force Cambridge Research Center, and shows that equal velocities of the two polarization waves can be achieved.

The present paper, which is intended to supplement existing literature, gives design equations and curves for the double-layer dielectric slab on a perfectly conducting surface. First, the character of the fields is studied, and then the conditions for circular polarization are investigated.

(2) THE TWO-LAYER EARTHED DIELECTRIC SLAB

The structure of the 2-layer earthed dielectric slab and its co-ordinate system are depicted in Fig. 1. All media are assumed
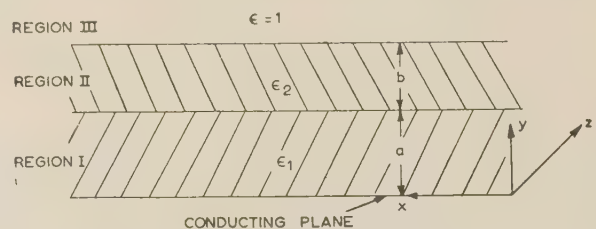
Fig. 1.—Two-layer earthed dielectric slab.

to be lossless, and a harmonic time dependence is assumed and suppressed. Rationalized M.K.S. units are used throughout. In the x- and z-directions the structure is unlimited, and on this lossless structure uncoupled modes can exist. Surface waves having no x-variation and propagating as $\exp(-j\beta z)$ in the z-direction are of interest.

(2.1) TM Waves

By the use of Hertzian vectors the field expressions can be immediately written as

$$H = \text{curl } u_z\psi, \quad E = \frac{1}{j\omega\epsilon} \text{curl curl } u_z\psi \quad . \quad . \quad (1)$$

where u_z is a unit vector in the z-direction and ψ is a scalar

---

* By arbitrary polarization is meant the combination of the two principal polarization components with arbitrary amplitudes and phase.

---

function that satisfies the scalar Helmholtz equation and the boundary conditions. The field has three components, given by

$$E_y = \frac{1}{j\omega\epsilon} \frac{\partial^2\psi}{\partial_y\partial_z}$$

$$E_z = \frac{-1}{j\omega\epsilon} \frac{\partial^2\psi}{\partial y^2}$$

$$H_x = \frac{\partial\psi}{\partial y}$$

$$\quad\quad . \quad . \quad . \quad . \quad (2)$$

In region I the longitudinal E must vanish at $y = 0$; therefore, if γ is the transverse wave number and A the amplitude, then

$$\psi_1 = A \begin{Bmatrix} \sin \\ \sinh \end{Bmatrix} \gamma y \; \exp(-j\beta z), \quad \epsilon_1 k^2 = \beta^2 \pm \gamma^2 \quad . \quad (3)$$

$$E_y = \frac{-\beta\gamma A}{\omega\epsilon_0\epsilon_1} \begin{Bmatrix} \cos \\ \cosh \end{Bmatrix} \gamma y \; \exp(-j\beta z)$$

$$E_z = \frac{\pm\gamma^2 A}{j\omega\epsilon_0\epsilon_1} \begin{Bmatrix} \sin \\ \sinh \end{Bmatrix} \gamma y \; \exp(-j\beta z) \quad . \quad . \quad (4)$$

$$H_x = \gamma A \begin{Bmatrix} \cos \\ \cosh \end{Bmatrix} \gamma y \; \exp(-j\beta z)$$

where the notation of curly brackets means that the transverse variation of the fields can be either trigonometric or hyperbolic; the upper sign is to be taken with the upper function in each case. All wave numbers are defined to be real and non-negative. Although all relative permeabilities are taken as unity, it is a trivial extension to include a value of μ for each layer.

If, for region II (the upper slab), the amplitude is B, then the components are

$$\psi_2 = B \begin{Bmatrix} \sin \\ \sinh \end{Bmatrix} (\delta y - C) \; \exp(-j\beta z), \quad \epsilon_2 k^2 = \beta^2 \pm \delta^2 \quad . \quad (5)$$

$$E_y = \frac{-\beta\delta B}{\omega\epsilon_0\epsilon_2} \begin{Bmatrix} \cos \\ \cosh \end{Bmatrix} (\delta y - C) \; \exp(-j\beta z)$$

$$E_z = \frac{\pm\delta^2 B}{j\omega\epsilon_0\epsilon_2} \begin{Bmatrix} \sin \\ \sinh \end{Bmatrix} (\delta y - C) \; \exp(-j\beta z) \quad . \quad (6)$$

$$H_x = \delta B \begin{Bmatrix} \cos \\ \cosh \end{Bmatrix} (\delta y - C) \; \exp(-j\beta z)$$

Here the transverse wave number is δ, and C is a constant. Again, the transverse variation may be either trigonometric or hyperbolic.

In region III, the air region, the wave must be attenuated as $\exp(-\alpha y)$; hence, by using unit amplitude one obtains

$$\psi_3 = \exp(-\alpha y) \exp(-j\beta z), \quad k^2 = \beta^2 - \alpha^2. \quad . \quad (7)$$

$$E_y = (\alpha\beta/\omega\epsilon_0) \exp(-\alpha y) \exp(-j\beta z)$$

$$E_z = (-\alpha^2/j\omega\epsilon_0) \exp(-\alpha y) \exp(-j\beta z) \quad . \quad . \quad (8)$$

$$H_x = -\alpha \exp(-\alpha y) \exp(-j\beta z)$$

At the boundaries $y = a$ and $y = a + b$, the tangential field components (or impedances) must match; this results in one equation independent of amplitudes at each boundary. These equations are

$$\alpha = \frac{\pm\delta}{\epsilon_2} \begin{Bmatrix} \tan \\ \tanh \end{Bmatrix} (\delta a + \delta b - C) \quad . \quad . \quad (9)$$

$$\frac{\pm\delta}{\epsilon_2} \begin{Bmatrix} \tan \\ \tanh \end{Bmatrix} (\delta a - C) = \frac{\pm\gamma}{\epsilon_1} \begin{Bmatrix} \tan \\ \tanh \end{Bmatrix} \gamma a \quad . \quad . \quad (10)$$

Note that if both hyperbolic functions are chosen, it is not possible to satisfy both (9) and (10) because (9) would then require $C > \delta(a + b)$, whereas (10) would require $C < \delta a$. Hence, the transverse variation must be trigonometric in at least one of the two slabs. Accordingly, three types of fields are possible and are defined as follows.

Type I.—This field has trigonometric transverse variation in both slabs. The set of Pythagorean relations for this case becomes

$$k^2 = \beta^2 - \alpha^2, \quad \epsilon_1 k^2 = \beta^2 + \gamma^2, \quad \epsilon_2 k^2 = \beta^2 + \delta^2 \quad . \quad (11)$$

From this it is clear that

$$\left. \begin{aligned} 1 \leqslant \beta/k \leqslant \sqrt{\epsilon_1} \\ 1 \leqslant \beta/k \leqslant \sqrt{\epsilon_2} \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad (12)$$

so that the reciprocal of the velocity ratio is limited to the square-root of the lesser dielectric constant. The constant C may be eliminated from (9) and (10) to give one equation independent of amplitudes:

$$\delta b = m\pi + \text{arc tan}\,(\epsilon_2\alpha/\delta) - \text{arc tan}\,[(\epsilon_2\gamma/\epsilon_1\delta)\tan\gamma a]$$
$$m = 0, 1, 2 \ldots \quad . \quad (13)$$

All of the arc-tangent functions used here and in subsequent equations will be principal values defined by

$$0 \leqslant \text{arc tan}\,X < \pi$$

The $m\pi$ term allows higher-order modes to be considered, the dominant mode being that for $m = 0$. Although eqn. (13) is written in terms of the transverse wave numbers α, γ and δ, it can be easily restated in terms of the ratio β/k by use of eqns. (11).

Type II.—This field has trigonometric variation in region II and hyperbolic variation in region I. The Pythagorean relations for type II furnish the following range for β/k:

$$\sqrt{\epsilon_1} \leqslant \beta/k \leqslant \sqrt{\epsilon_2} . \quad . \quad . \quad . \quad (14)$$

This type II field exists only when $\epsilon_2 > \epsilon_1$. Eqns. (9) and (10) give as the determining equation

$$\delta b = m\pi + \text{arc tan}\,(\epsilon_2\alpha/\delta) + \text{arc tan}\,[(\epsilon_2\gamma/\epsilon_1\delta)\tanh\gamma a]$$
$$m = 0, 1, 2 \ldots \quad . \quad (15)$$

These fields are just a continuation of the type I fields for which $\epsilon_2 > \epsilon_1$; this will be made clear later.

Type III.—This field is just the reverse of the preceding case, i.e. hyperbolic variation in the upper slab and a trigonometric variation in the slab adjacent to the earth plane. The pertinent equations for this field imply the range

$$\sqrt{\epsilon_2} \leqslant \beta/k \leqslant \sqrt{\epsilon_1} . \quad . \quad . \quad . \quad (16)$$

so that the system must have $\epsilon_1 > \epsilon_2$. The determining equation for type III fields is

$$\delta b = \text{arc tanh}\,[(\epsilon_2\gamma/\epsilon_1\delta)\tan\gamma a] - \text{arc tanh}\,(\epsilon_2\alpha/\delta) \quad . \quad (17)$$

Since all the wave numbers are defined to be positive, the argument γa in eqn. (17) must be restricted to the range $0 < \gamma a < \pi/2$; otherwise a negative δb would result.

The results of calculations made by using eqns. (13), (15) and (17) are presented graphically for several combinations of dielectric constants in the two layers. Fig. 2 shows β/k for the first two type I fields ($m = 0$ and $m = 1$) and, in addition, the lowest type II field for $\epsilon_1 = 2$ and $\epsilon_2 = 4$. Primes on β/k denote the second mode ($m = 1$) in all figures. As the β/k ratio goes through the value $\sqrt{\epsilon_1}$ the field changes (in mathematical description) from type I to type II; the type II field is shown by broken
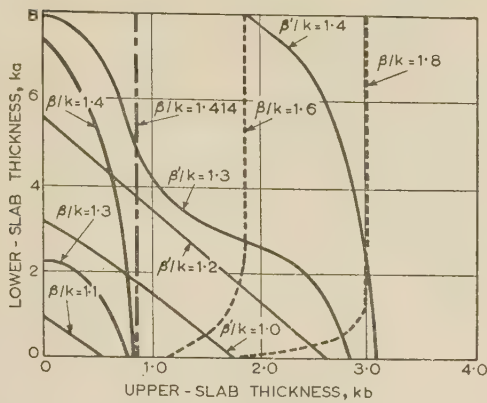
Fig. 2.—Inverse velocity ratios for TM waves for field types I and II.
$\epsilon_1 = 2$, $\epsilon_2 = 4$.
Type I ———.   Type II – – –.   Types I and II — –.
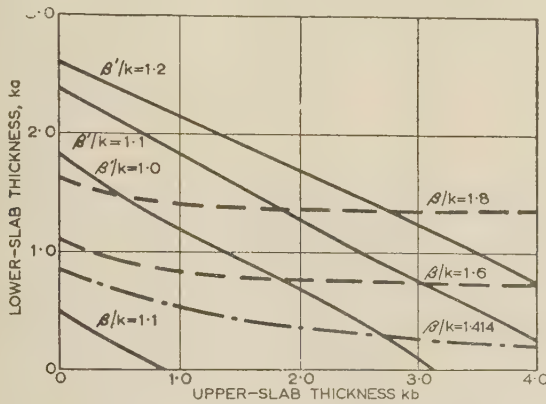
Fig. 3.—Inverse velocity ratios for TM waves for field types I and III.
$\epsilon_1 = 4$, $\epsilon_2 = 2$.
Type I ———.   Type III – – –.   Types I and III — –.

lines in the diagrams.   In Fig. 3 similar curves are plotted for $\epsilon_1 = 4$ and $\epsilon_2 = 2$, except that the lowest type I field becomes a type III field for $\beta/k > \sqrt{\epsilon_2}$.   Fig. 4 shows curves of the type II field for an upper slab of polystyrene and a lower slab of air. Additional large-size graphs are given in a report[8] of which this is a summary.
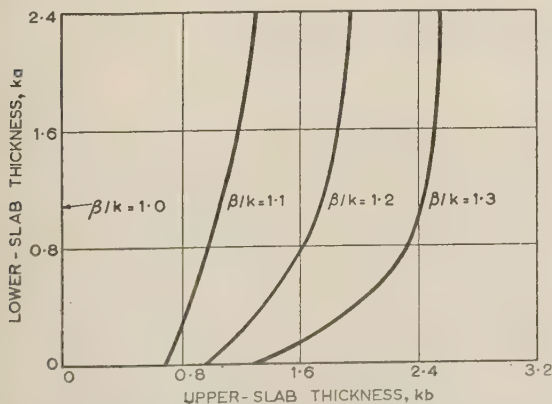
Fig. 4.—Inverse velocity ratios for TM waves for type II.
$\epsilon_1 = 1$, $\epsilon_2 = 2 \cdot 5$.

(2.2) TE Waves

The mathematical development for these waves is exactly analogous to that for the TM waves; starting with Hertzian vectors, one writes the field in each of three regions, matches tangential field components at the two interfaces, simplifies the resulting equations to get the final equation for the propagation coefficient, and graphs the solutions for several cases.   The symbols employed are those used with the TM waves, although each has a new value in the discussion that follows.   The key equations alone are given, and the various field expressions will follow.   The field as obtained from the Hertzian vector is written:

$$E = \text{curl } u_z \phi, \quad H = (-1/j\omega\mu) \text{ curl curl } u_z\phi \quad . \quad (18)$$

The three scalar functions are:

$$\phi_1 = A \begin{Bmatrix} \cos \\ \cosh \end{Bmatrix} \gamma y \} \exp(-j\beta z)$$
$$\phi_2 = B \begin{Bmatrix} \cos \\ \cosh \end{Bmatrix} (\delta y - D) \} \exp(-j\beta z) \quad \} \quad . \quad . \quad (19)$$
$$\phi_3 = \exp(-\alpha y) \exp(-j\beta z)$$

Once again the transverse impedances when matched at the boundaries $y = a$ and $y = a + b$, give a pair of equations analogous to (9) and (10); the new pair is

$$(-1/\alpha) = \frac{1}{\delta}\begin{Bmatrix} \tan \\ \tanh \end{Bmatrix} (\delta a + \delta b - D) \} \quad . \quad . \quad (20)$$

and

$$\frac{1}{\delta}\begin{Bmatrix} \tan \\ \tanh \end{Bmatrix} (\delta a - D) \} = \frac{1}{\gamma}\begin{Bmatrix} \tan \\ \tanh \end{Bmatrix} \gamma a \} \quad . \quad . \quad . \quad (21)$$

Here, as in the equations obtained for the TM modes, at least one region must have a trigonometric transverse variation; this gives the same three field types identified previously.   For each case, the Pythagorean relation (hence the range of β/k) is exactly the same as in the corresponding TM case.   Only the determining equations are new, and these are listed below.

Type I.
$$\delta b = n\pi - \arctan(\delta/\alpha) - \arctan[(\delta/\gamma)\tan\gamma a]$$
$$n = 1, 2, 3 \ldots$$
$$. \quad . \quad . \quad . \quad (22)$$

Note that the lowest mode obtains for $n = 1$, which indicates a cut-off frequency for all TE waves.

Type II.
$$\delta b = n\pi - \arctan(\delta/\alpha) - \arctan[(\delta/\gamma)\tanh\gamma a]$$
$$n = 1, 2, 3 \ldots$$
$$. \quad . \quad . \quad . \quad (23)$$

Type III.
$$\delta b = - \text{arc tanh}(\delta/\alpha) - \text{arc tanh}[(\delta/\gamma)\tan\gamma a] \quad . \quad (24)$$

In the type III fields the argument γa must be between $\pi/2$ and π in order that eqn. (24) may be satisfied.   Figs. 5 and 6 show the results of calculations using eqns. (22), (23), and (24) for the same combinations of dielectric constants used in the TM plots.   Fig. 7 is a graph of the TE type II mode for $\epsilon_2 = 2 \cdot 5$ and $\epsilon_1 = 1 \cdot 0$.

The graphical data portray the manner in which the field parameters vary with changes in dielectric constants and slab thicknesses.   The distribution of power in the transverse direction is of interest; in general, the power is concentrated in the

vicinity of the slab that has the trigonometric variation in the y-direction. A crude estimate of the amount of power flow on the free-space side of this slab is simply that which occurs outside of a single-slab structure, the single slab having the same
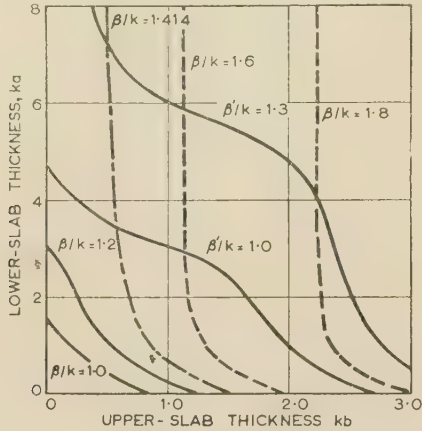
Fig. 5.—Inverse velocity ratios for TE waves for field types I and II.
$\epsilon_1 = 2$, $\epsilon_2 = 4$.
Type I ———. Type II – – –. Types I and II — –.
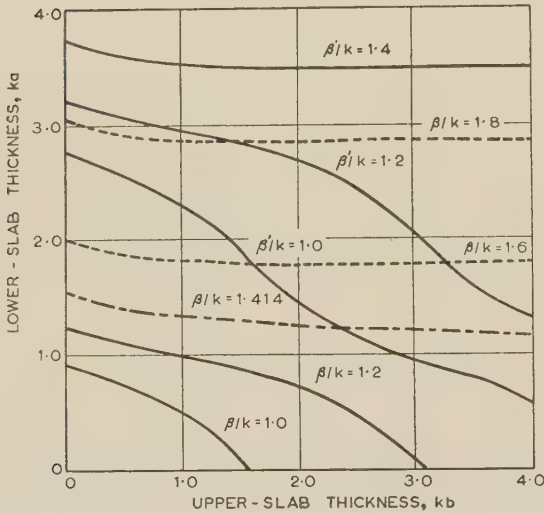
Fig. 6.—Inverse velocity ratios for TE waves for field types I and III.
$\epsilon_1 = 4$, $\epsilon_2 = 2$.
Type I ———. Type III – – –. Types I and III — –.
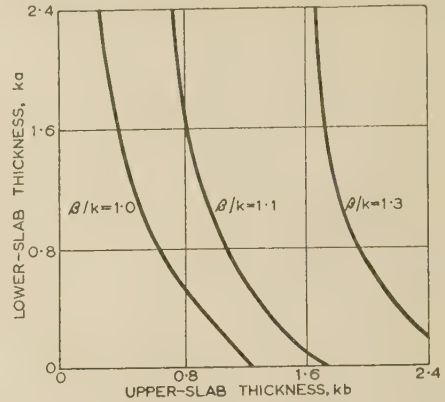
Fig. 7.—Inverse velocity ratios for TE waves for type II.
$\epsilon_1 = 1$, $\epsilon_2 = 2 \cdot 5$.

dielectric constant and thickness as the slab referred to above. Reference could also be made to the power-flow curves (for $\epsilon_2 = 2 \cdot 5$ and ϵ_1 near unity) of Jones and Cohn.[3]

Before proceeding to a consideration of the conditions for arbitrary polarization of surface waves over the double-slab system, it is useful to summarize the results of the analysis thus far.

The analysis has shown that there are three types of field for each polarization and that these types are identified by the nature of the variation in the direction normal to the surface. These classifications, together with the corresponding conditions on the parameters, are reviewed in Table 1. It is instructive to discuss the field types on the basis of the two possible arrangements of the dielectric constants in the two layers: $\epsilon_2 > \epsilon_1$ and $\epsilon_1 > \epsilon_2$. The type I fields are common to both of these combinations, whereas type II fields exist with only the former arrangement, and type III fields are possible with only the latter. For a configuration having $\epsilon_2 > \epsilon_1$, both type I and type II fields are possible. Type I exists exclusively at the lower frequencies, and both types exist simultaneously as the frequency is increased beyond that for which $\beta/k = \sqrt{\epsilon_1}$ for the dominant type I field. On the other hand, if $\epsilon_2 < \epsilon_1$, then both type I and type III fields are possible; again the former exists exclusively at lower frequencies, and both types exist simultaneously at frequencies in excess of that for which $\beta/k = \sqrt{\epsilon_2}$ for the dominant type I field. These remarks apply to both TM and TE polarizations. In fact, if the slab parameters and frequency are such that the TE polarization can propagate, then both polarizations are possible, although, in general, the velocities of the two polarizations will differ markedly except under certain specific conditions as established in the following Section.

Table 1

FIELD CLASSIFICATIONS AND CONDITIONS ON THE PARAMETERS

Mathematical classification of modes	Field variation in the direction normal to the slab		Arrangement of dielectric constants required	Range of the inverse velocity ratio
	Region I	Region II		
Type I	Trigonometric ..	Trigonometric ..	$\epsilon_2 > \epsilon_1$ or $\epsilon_2 < \epsilon_1$	$1 \leqslant \beta/k \leqslant \sqrt{\epsilon_1}$ or $1 \leqslant \beta/k \leqslant \sqrt{\epsilon_2}$
Type II	Hyperbolic	Trigonometric ..	$\epsilon_2 > \epsilon_1$	$\sqrt{\epsilon_1} \leqslant \beta/k \leqslant \sqrt{\epsilon_2}$
Type III	Trigonometric ..	Hyperbolic	$\epsilon_2 < \epsilon_1$	$\sqrt{\epsilon_2} \leqslant \beta/k \leqslant \sqrt{\epsilon_1}$

(3) CONDITIONS FOR THE POSSIBILITY OF PROPAGATING ARBITRARY POLARIZATION

Because the mathematical descriptions of the double-slab fields fall into three categories, it is useful to consider the possibilities of making the propagation coefficients of TM and TE waves equal for each of the three types of fields for particular combinations of dielectric constants and slab thicknesses. These types are treated in inverse order, starting with type III.

If the two principal polarizations (TM and TE waves) travel with the same phase velocity ω/β, then α, γ, and δ are identical for the two waves. But in order for TM type III to exist, $0 \leqslant \gamma a < \pi/2$, whereas for the TE case $\pi/2 \leqslant \gamma a < \pi$. Hence, it is immediately clear that arbitrary or circular polarization is not possible with type III fields.

Type II fields enjoy the complementary position, i.e. circular polarization is always possible. This can be shown by equating the right-hand sides of eqns. (15) and (23), which gives

$$(n - m)\pi = \arctan(\delta/\alpha) + \arctan(\epsilon_2\alpha/\delta)$$
$$+ \arctan[(\delta/\gamma)\tanh \gamma a] + \arctan[(\epsilon_2\gamma/\epsilon_1\delta)\tanh \gamma a] . \quad (25)$$

Each angle is less than $\pi/2$, so that at the outset n is limited to $n = m + 1$ and the modes must 'pair up'. Now let $\tanh \gamma a = T$, and express eqn. (25) in terms of T. This results in a quadratic equation in T:

$$T^2\epsilon_2\alpha + T\alpha(\epsilon_2 - \epsilon_1) - \gamma\epsilon_1 = 0 \quad . \quad . \quad . \quad (26)$$

Since the discriminant is always positive, there are two real and unequal roots; furthermore, only one value of T is positive, namely

$$T = \frac{[\alpha^2(\epsilon_2 - \epsilon_1)^2 + 4\gamma^2\epsilon_1\epsilon_2]^{1/2} - \alpha(\epsilon_2 - \epsilon_1)}{2\epsilon_2\gamma} \quad . \quad (27)$$

Thus, for a given ϵ_1, ϵ_2 and velocity ω/β, there is only one set of thicknesses ka and kb that allows arbitrary polarization. From eqn. (27) the value of ka is immediately

$$ka = (k/\gamma) \operatorname{arc\,tanh} T \quad . \quad . \quad . \quad . \quad (28)$$

and the value kb, from eqn. (15), is

$$kb = (k/\delta)[m\pi + \arctan(\epsilon_2\alpha/\delta) + \arctan(\epsilon_2\gamma T/\epsilon_1\delta)] \quad . \quad (29)$$

To recapitulate: if the dielectric constants are known, the two thicknesses necessary to produce a desired β/k ratio can be readily found from eqns. (28) and (29). Some numbers have been plotted in Fig. 8, giving the variation of β/k with ka and kb for $\epsilon_1 = 1$ and for several values of ϵ_2. One family of curves in this graph is for constant β/k; the second family contains a curve for each set of dielectric constants. The diagram gives an idea of the interrelation of thicknesses and dielectric constants for a given velocity, or the interrelation of thicknesses and velocity for a given pair of dielectric constants that is required for arbitrary polarization.
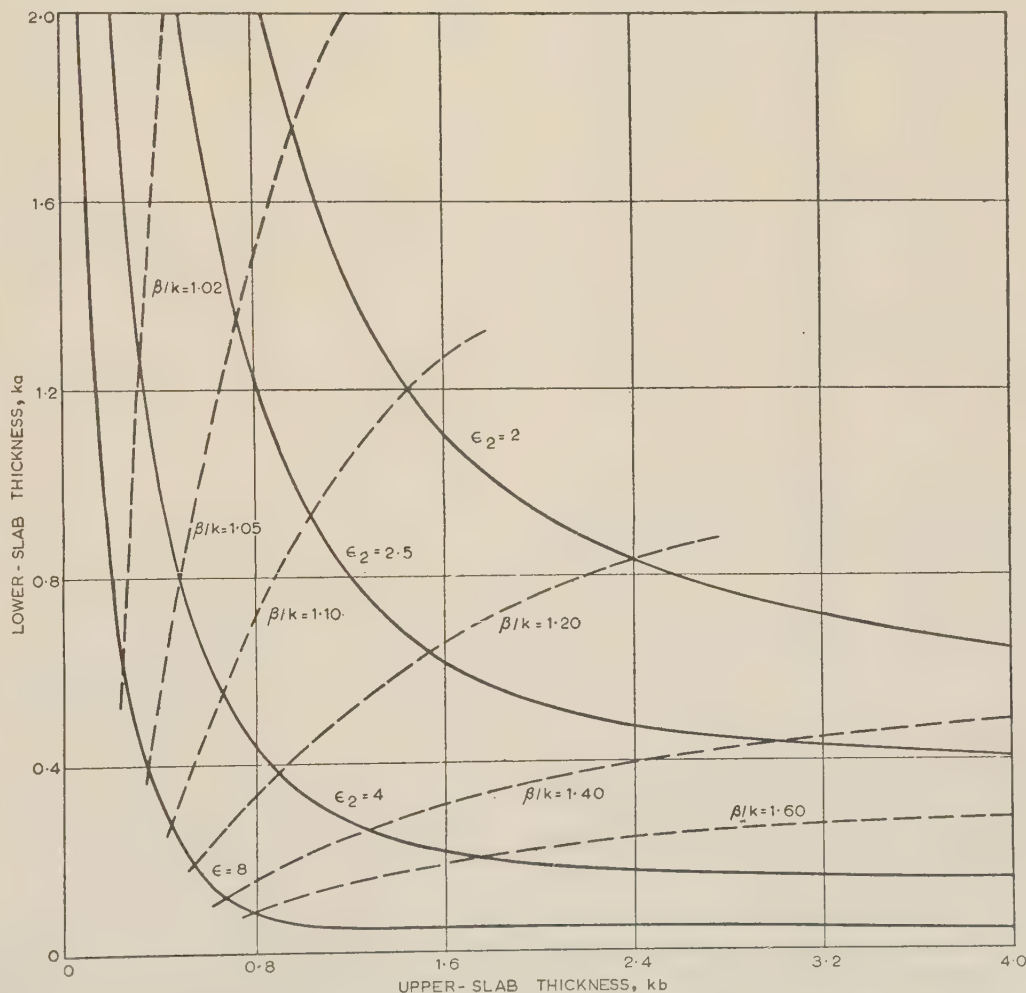
Fig. 8.—Contours of equal phase velocity for TM and TE waves on the slab thickness plane for field type II.

$\epsilon_1 = 1 \cdot 0$.

Inverse velocity — —. $\epsilon_2 = $ Constant ——.

Type I fields are more complex in that circular polarization is only sometimes possible. To develop this idea, equate the right-hand sides of eqns. (13) and (22); this gives

$$(n - m)\pi = \arctan(\epsilon_2\alpha/\delta) + \arctan(\delta/\alpha)$$
$$+ \arctan[(\delta/\gamma)\tan\gamma a] - \arctan[(\epsilon_2\gamma/\epsilon_1\delta)\tan\gamma a] \quad . \quad (30)$$

Again $n = m + 1$, i.e. the modes go in pairs. Now let $\tan\gamma a = T$, and rearrange eqn. (30) into a quadratic equation in T:

$$T^2\epsilon_2\gamma - T\alpha(\epsilon_2 - \epsilon_1) + \epsilon_1\gamma = 0 \quad . \quad . \quad . \quad (31)$$

To have real roots, the discriminant must be greater than zero. This gives the condition

$$\alpha^2(\epsilon_2 - \epsilon_1)^2 - 4\epsilon_1\epsilon_2\gamma^2 \geqslant 0$$

which, using eqns. (11), reduces to a condition on β/k:

$$(\beta/k)^2 \geqslant 1 + 4\epsilon_1\epsilon_2(\epsilon_1 - 1)/(\epsilon_1 + \epsilon_2)^2 \quad . \quad . \quad . \quad (32)$$

This is a severe restriction on β/k as will be seen below. The two real roots are, of course, equal if the inequality of (32) becomes an equality.

Values of T are thus

$$T = \frac{\alpha(\epsilon_2 - \epsilon_1) \pm [\alpha^2(\epsilon_2 - \epsilon_1)^2 - 4\epsilon_1\epsilon_2\gamma^2]^{1/2}}{2\epsilon_2\gamma} \quad . \quad (33)$$

so that to a particular set of dielectric constants and a particular β/k there correspond two sets of thicknesses ka and kb that give the β/k for both polarizations. Note that both values of T are positive. Again, ka and kb are immediately found from the equations

$$ka = (k/\gamma)\arctan T \quad . \quad . \quad . \quad . \quad (34)$$

and

$$kb = (k/\delta)[m\pi + \arctan(\epsilon_2\alpha/\delta) - \arctan(\epsilon_2\gamma T/\epsilon_1\delta)] \quad . \quad (35)$$

These thicknesses are plotted for various values of β/k in Figs. 9(a) and 9(b) for dielectric-constant sets of ($\epsilon_1 = 1 \cdot 25$, $\epsilon_2 = 2 \cdot 5$), and ($\epsilon_1 = 1 \cdot 2$, $\epsilon_2 = 12$), respectively. To take a numerical example of the limitation of β/k imposed by (32), choose $\epsilon_1 = 2$ and $\epsilon_2 = 4$; this gives

$$1\tfrac{8}{9} \leqslant (\beta/k)^2 < 2$$

which is a very narrow range for the inverse velocity ratio.

Other restrictions on the type I case arise from eqn. (30). Here the sum of three angles minus a fourth angle must equal π. If the negative angle is greater than any one of the other three, this sum cannot add to π, since $\tan\gamma a > 0$. So we require that

$$\left.\begin{array}{l} (\epsilon_2\gamma/\epsilon_1\delta)\tan\gamma a < (\epsilon_2\alpha/\delta) \\ \qquad\qquad < (\delta/\alpha) \\ \qquad\qquad < (\delta/\gamma)\tan\gamma a \end{array}\right\} \quad . \quad . \quad . \quad (36)$$

One of these inequalities will reduce to the condition $\epsilon_2 > \epsilon_1$. The important fact is that $\epsilon_2 > \epsilon_1$ is always a requirement for arbitrary polarization, since the type II modes exist only for $\epsilon_2 > \epsilon_1$. Thus, the top slab must always have a greater dielectric constant than the lower slab to allow the velocities to be made equal.

The equivelocity points displayed in Figs. 8 and 9 are, of course, valid at only a single frequency. For fixed slab thicknesses a and b it is readily seen that the velocities for both TM and TE waves are decreasing functions of frequency with the TE wave velocity varying the most rapidly. The actual functional dependence for each wave can be determined only by numerical

(a)

(b)
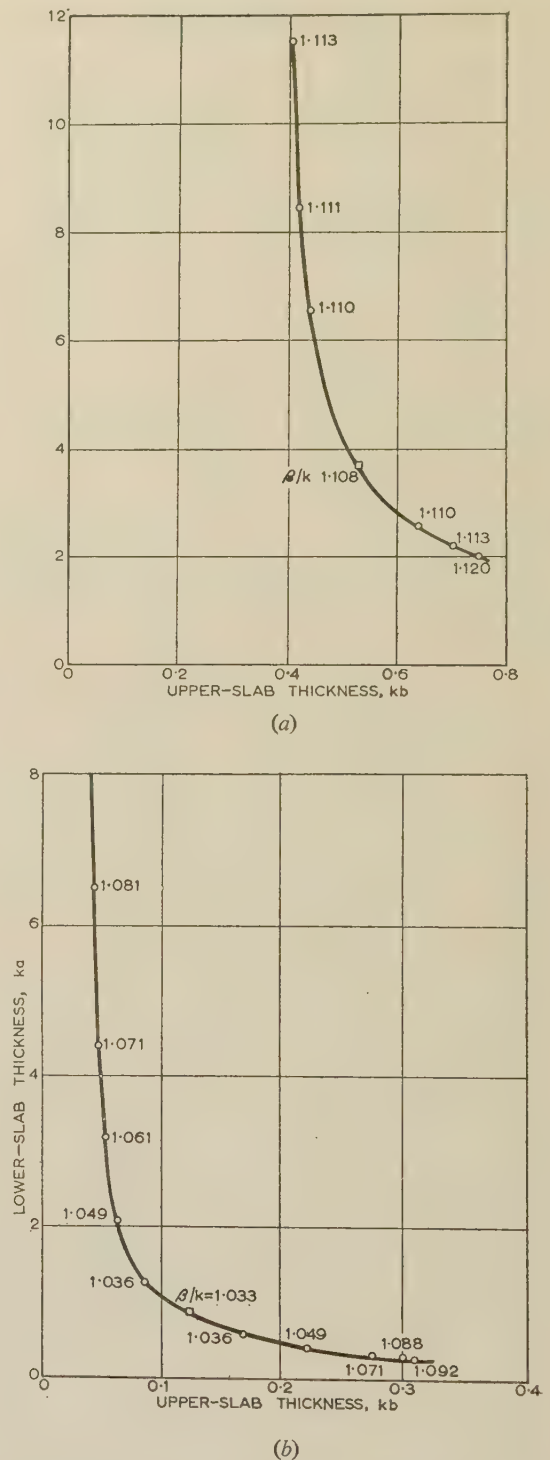
Fig. 9.—Equivelocity distribution on the slab-thickness plane for type I.

(a) $\epsilon_1 = 1 \cdot 25$, $\epsilon_2 = 2 \cdot 5$.
(b) $\epsilon_1 = 1 \cdot 2$, $\epsilon_2 = 12 \cdot 0$.

means. Hence, an evaluation of the amount of the velocity difference between the two waves when the frequency deviates from the equivelocity value is most readily accomplished by recourse to the graphical data. It is observed, for example, that the least velocity separation occurs for a given frequency change when the lower slab thickness is chosen greater than the upper

slab thickness. To illustrate this fact, choose $\epsilon_1 = 1$ and $\epsilon_2 = 2 \cdot 5$ and observe that for $a/b \simeq 4 \cdot 5$ ($\beta/k = 1 \cdot 02$), a 5% frequency change produces approximately $0 \cdot 5\%$ velocity difference, whereas for $a/b \simeq 0 \cdot 4$ ($\beta/k = 1 \cdot 20$), the same frequency change produces approximately 2% velocity difference.

The results of the foregoing and additional examples involving different combinations of dielectric constants indicate that the frequency sensitivity of the double-slab system is generally rather mild. The observations that have been made are quite broad. The designer will glean more specific information relative to a particular system by careful examination of the appropriate curves.

(4) CONCLUSIONS

The fields in a 2-layer earthed dielectric slab must have a trigonometric transverse variation in at least one of the layers. For circular or arbitrary polarization to be admissible, the upper layer must have a greater dielectric constant than the lower. For arbitrary polarization, the inverse velocity ratio β/k is restricted to the range

$$\left[\epsilon_1 - \frac{(\epsilon_1 - 1)(\epsilon_2 - \epsilon_1)^2}{(\epsilon_1 + \epsilon_2)^2} \right]^{1/2} \leqslant \beta/k \leqslant \sqrt{\epsilon_2}$$

which is just slightly greater than the range

$$\sqrt{\epsilon_1} \leqslant \beta/k \leqslant \sqrt{\epsilon_2}$$

Design equations and data are presented that allow a structure to be designed for equal phase velocities of TM and TE waves and that indicate how the several parameters interact. In the narrow range (of β/k) between $\left[\epsilon_1 - \frac{(\epsilon_1 - 1)(\epsilon_2 - \epsilon_1)^2}{(\epsilon_1 + \epsilon_2)^2} \right]^{1/2}$ and $\sqrt{\epsilon_1}$, dual solutions exist, i.e. for a given pair of dielectric constants, two sets of thicknesses give the same velocity; in the remainder of the range the solution is unique.

(6) REFERENCES

(1) HANSEN, ROBERT C.: 'Single-Slab Circular-Polarization Surface-Wave Structure', Scientific Report No. 9, Research Laboratories, Hughes Aircraft Company, February, 1956. To appear in *IRE Transactions*, **MTT-5**, April, 1957.

(2) KARBOWIAK, A. E.: 'Theory of Composite Guides: Stratified Guides for Surface Waves', *Proceedings I.E.E.*, Paper No. 1659 R, July, 1954 (**101**, Part III, p. 238).

(3) EHRENSPECK, H., GERBES, W., and ZUCKER, F. J.: 'Trapped Wave Antennas', *Convention Record of the Institute of Radio Engineers*, 1954, **2**, Part I, p. 25.

(4) JONES, E. M. T., and COHN, S. B.: 'A Study of Radiating Structures for Perpendicularly-Polarized Flush Radar Antennas', Contract No. AF19(604)–266, Stanford Research Institute Technical Report 44, July, 1954.

(5) KAY, A. F.: 'The Excitation of Surface Waves in Multi-Layered Media', Contract No. AF19(604)–1126, Technical Research Group, 1954.

(6) KAY, A. F., and ZUCKER, F. J.: 'Efficiency of Surface Waves Excitation', *Convention Record of the Institute of Radio Engineers*, 1955, **3**, Part I, p. 1.

(7) ZUCKER, FRANCIS J.: 'The Guiding and Radiation of Surface Waves', *Proceedings of the Symposium on Modern Advances in Microwave Techniques*, Polytechnic Institute of Brooklyn, November, 1954, p. 403.

(8) PLUMMER, ROBERT E., and HANSEN, ROBERT C.: 'Double-Slab Arbitrary-Polarization Surface-Wave Structure', Scientific Report No. 10, Research Laboratories, Hughes Aircraft Company, June, 1956.

621.372.8 : 538.56

A NOTE ON THE EXCITATION OF SURFACE WAVES

By Prof. A. L. CULLEN, Ph.D., Associate Member.

SUMMARY

The paper gives an extension of results on the launching efficiency of surface waves obtained in a previous paper.

In particular, a study is made of the effect on the launching efficiency curves of varying the ratio k_1/k, which is a measure of the surface reactance of the guiding surface.

LIST OF PRINCIPAL SYMBOLS

E_y = y-component of electric field strength.
H_z = z-component of magnetic field strength.
H = Height of aperture.
h = Height of source slot above guiding surface.
k = Phase constant for free space.
k_1 = Decay constant of surface wave.
$V(h)$ = Voltage across slot at height h.
x, y, z = Rectangular co-ordinates.
β = Phase-change coefficient in x-direction.
β_1 = Phase-change coefficient of surface wave.
η = Launching efficiency.
ϵ_0 = Permittivity of free space.
λ = Free-space wavelength.
μ_0 = Permeability of free space.

(1) INTRODUCTION

In an earlier paper the excitation of surface waves was discussed. The surface-wave launching efficiency of a 'chopped surface-wave' aperture distribution was considered and plotted as a function of aperture height for a particular value of decay constant k_1, namely $k_1 = 0 \cdot 5k$.

The purpose of the paper is to extend the previous results to other values of k_1, and to present formulae by means of which this extension can be carried out more easily than by the method of numerical integration previously employed.

The opportunity is taken to show that integration of the Poynting vector over the aperture plane itself can conveniently be used in the derivation of the formulae needed if we use a different path of integration from the usual Sommerfeld contour.

(2) CALCULATION OF APERTURE MAGNETIC FIELD

In a previous paper,[1] the following two formulae were derived for the magnetic field H_z at the point (x, y) produced by a voltage V across a narrow slot in a perfectly conducting vertical screen, the slot being situated at a height h above a corrugated guiding surface:

$$H_{z1} = -j\frac{k}{2\pi}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}V\int_{-\infty}^{+\infty}\frac{\varepsilon^{-uy}}{u}\left(\varepsilon^{uh} + \frac{u+k_1}{u-k_1}\varepsilon^{-uh}\right)\varepsilon^{-j\beta x}d\beta \quad (1)$$

or $H_{z1} = Vf(y, h)$ for $y \geqslant h$

$$H_{z2} = -j\frac{k}{2\pi}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}V\int_{-\infty}^{+\infty}\frac{\varepsilon^{-uh}}{u}\left(\varepsilon^{uy} + \frac{u+k_1}{u-k_1}\varepsilon^{-uy}\right)\varepsilon^{-j\beta x}d\beta \quad (2)$$

or $H_{z2} = Vf(h, y)$ for $0 \leqslant y < h$

where

$$u = \sqrt{(\beta^2 - k^2)} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

As shown elsewhere,[1] it is possible to generalize eqns. (1) and (2) to apply to any specified vertically polarized aperture distribution, by regarding the aperture electric field as being produced by a continuous distribution of 'slots' each with the appropriate exciting voltage. Mathematically speaking, we regard eqns. (1) and (2) as giving Green's function for the problem.
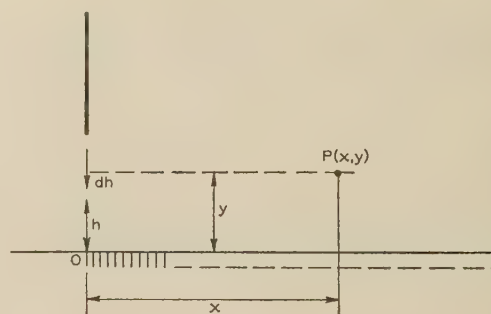
Fig. 1.—Illustrating the notation.

For the case of a 'chopped surface-wave' aperture distribution we therefore put, in the notation of Fig. 1,

$$\left.\begin{array}{l} V(h) = -E_0\varepsilon^{-k_1h}dh; \quad 0 < h \leqslant H \\ V(h) = 0; \quad h > H \end{array}\right\} \quad . \quad . \quad (4)$$

To find the field at a height y less than h we must bear in mind that for the 'slots' below y eqn. (1) must be used and for the 'slots' above y eqn. (2) must be used.

Thus, the total magnetic field H_z at the point (x, y) due to the aperture distribution specified by eqns. (4) is given by

$$H_z = \int_0^y V(h)f(y, h)dh + \int_y^H V(h)f(h, y)dh \quad . \quad . \quad (5)$$

Substituting eqns. (4) in (1) and (2) and integrating with respect to h as indicated in eqn. (5), we get

$$\frac{2\pi}{jk}\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}H_z = E_0\varepsilon^{-k_1y}\int_{-\infty}^{+\infty}\left[\frac{1}{u(u-k_1)} + \frac{1}{u(u+k_1)}\right]\varepsilon^{-j\beta x}d\beta$$

$$- E_0\varepsilon^{-k_1H}\int_{-\infty}^{+\infty}\left[\frac{\varepsilon^{-u(H+y)}}{u(u-k_1)} + \frac{\varepsilon^{-u(H-y)}}{u(u+k_1)}\right]\varepsilon^{-j\beta x}d\beta \quad . \quad (6)$$

[472]

The first integral in eqn. (6) could be evaluated by complex integration, but it is simpler to proceed as follows. When $H = \infty$, the field to the right of the aperture plane is a pure surface wave, and

$$\frac{\beta_1}{k}\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}H_z = E_0 \varepsilon^{-k_1 y}\varepsilon^{-j\beta_1 x} \quad . \quad . \quad . \quad . \quad (7)$$

The second term in eqn. (6) vanishes when $H = \infty$ on account of the factor $\varepsilon^{-k_1 H}$, bearing in mind that u is either real and positive or imaginary along the real axis of the β-plane. Thus, the first integral of eqn. (6) must have the value

$$\frac{2\pi\varepsilon^{-j\beta_1 x}}{j\beta_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Since this integral is independent of H we have in general

$$\frac{\beta_1}{k}\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}H_z = E_0 \varepsilon^{-k_1 y}\varepsilon^{-j\beta_1 x}$$

$$+ \frac{\beta_1}{2\pi j}E_0 \varepsilon^{-k_1 H}\int_{-\infty}^{+\infty}\left[\frac{\varepsilon^{-u(H+y)}}{u(u-k_1)} + \frac{\varepsilon^{-u(H-y)}}{u(u+k_1)}\right]\varepsilon^{-j\beta x}d\beta \quad . \quad (9)$$

We now apply the method of contour integration to evaluate the integral in eqn. (9).
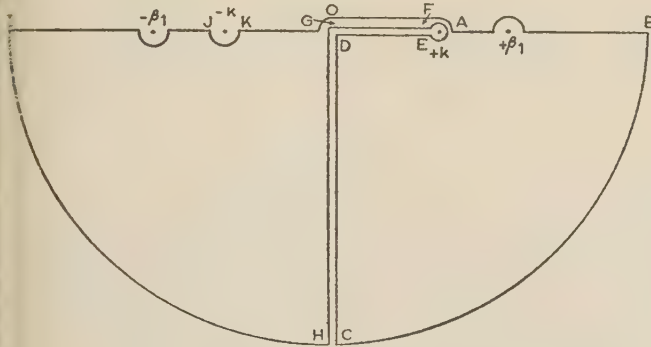
Fig. 2.—Contour of integration.

For the present purpose the contour shown in Fig. 2 is more convenient than that used in the previous paper, for the branch-cut contribution is now divided into two parts which give separately the active and reactive components of H_z. The active part of H_z is found from the integration along DE and FG; the integration along CD and GH would yield the reactive part, which is not needed. The pole at β_1 gives a contribution to the surface-wave field. The active part H_z' of H_z in the aperture plane ($x = 0$) is then given by

$$\frac{\beta_1}{k}\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)}H_z' = E_0 \varepsilon^{-k_1 y}(1 - \varepsilon^{-2k_1 H})$$

$$- \frac{\beta_1 E_0 \varepsilon^{-k_1 H}}{\pi}\int_0^k\left[\frac{k_1(\cos\psi - \cos\phi)}{\sqrt{(k^2 - \beta^2)(\beta^2 - \beta_1^2)}} + \frac{\sin\psi + \sin\phi}{(\beta^2 - \beta_1^2)}\right]d\beta$$

$$. \quad . \quad . \quad . \quad . \quad (10)$$

where
$$\psi = \sqrt{(k^2 - \beta^2)}(H + y) \left.\begin{array}{c}\\\\\end{array}\right\} \quad . \quad . \quad . \quad . \quad (11)$$
$$\phi = \sqrt{(k^2 - \beta^2)}(H - y)$$

(3) POWER FLOW

To calculate the power flowing through the aperture, we must multiply H_z, as given by eqn. (10), by

$$E_z = E_0 \varepsilon^{-ky} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

and integrate with respect to y from O to H. Assuming that it is permissible to reverse the order of integration, the integration with respect to y can be carried through for the second term in eqn. (10) as well as for the first. If we write

$$\Phi(y) = \int_0^k \frac{\cos\psi - \cos\phi}{\sqrt{(k^2 - \beta^2)(\beta^2 - \beta_1^2)}}d\beta \quad . \quad . \quad (13)$$

it is easily verified that

$$\frac{d\Phi(y)}{dy} = -\int_0^k \frac{\sin\psi + \sin\phi}{\beta^2 - \beta_1^2}d\beta \quad . \quad . \quad (14)$$

The integration of $E_0 \varepsilon^{-k_1 y}H_z'$ with respect to y can then be carried out by parts as follows:

$$\int_0^H \varepsilon^{-k_1 y}\left[k_1\Phi(y) - \frac{d\Phi(y)}{dy}\right]dy = -\left[\varepsilon^{-k_1 y}\Phi(y)\right]_0^H \quad . \quad (15)$$

Using eqns. (10), (13), (14) and (15) the total power flow through the aperture $0 \leqslant y \leqslant H$ is found to be

$$P = \sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}E_0^2 \frac{k}{2k_1\beta_1}(1 - \varepsilon^{-2k_1 H})^2$$

$$+ \sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}E_0^2 \varepsilon^{-2k_1 H}\frac{2k}{\pi}\int_0^k \frac{\sin^2\sqrt{(k^2 - \beta^2)}H}{\sqrt{(k^2 - \beta^2)(\beta^2 - \beta_1^2)}}d\beta \quad . \quad (16)$$

The first term of eqn. (16) is the power in the surface wave, and the substitution $\beta = k\cos\theta$ in the second term puts it into a form recognizable as the radiated power, thus:

$$P = \sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}E_0^2 \frac{k}{2k_1\beta_1}(1 - \varepsilon^{-2k_1 H})^2$$

$$+ \sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}E_0^2 \varepsilon^{-2k_1 H}\frac{2k}{\pi}\int_0^{\pi/2} \frac{\sin^2(kH\sin\theta)}{k_1^2 + k^2\sin^2\theta}d\theta \quad . \quad (17)$$

An infinite-series representation of the integral in eqn. (17) is obtained in Section 7, and hence for the ratio of radiated power in the surface wave we find

$$\frac{P_R}{P_S} = \frac{1 - J_0(2kH) - 2\sum_{m=1}^{\infty}J_{2m}(2kH)\left(\frac{\beta_1 - k_1}{k}\right)^{2m}}{2\sinh^2 k_1 H} \quad . \quad (18)$$

In the limiting case $2kH \to \infty$ this reduces to

$$\frac{P_R}{P_S} = \frac{1}{2\sinh^2 k_1 H} \quad . \quad . \quad . \quad . \quad (19)$$

It might be expected that the result for launching efficiency obtained by Rich[2] would lead to a value of P_R/P_S agreeing with eqn. (19) for very short wavelengths.

Rich's formula is

$$\eta = 1 - \varepsilon^{-2k_1 H} \quad . \quad . \quad . \quad . \quad (20)$$

in our notation, and the corresponding result for P_R/P_S can be written

$$\frac{P_R}{P_S} = \frac{1}{2\sinh^2 k_1 H}\left(\frac{1 - \varepsilon^{-2k_1 H}}{2}\right) \quad . \quad . \quad (21)$$

The discrepancy between Rich's result and the results given here have been discussed previously[3] for the special case

$k_1/k = 0.5$, but it is interesting to see that the present results do not agree with Rich's even if $k_1/k \to 0$. Rich relates his launching efficiency to the power carried by the incident plane wave rather than to the power flow through the aperture, and in Rich's view this accounts for the discrepancy in the case $k_1/k = 0.5$. However, even bearing in mind the different definition, it seems unlikely that the launching efficiency as defined by Rich would be independent of k_1/k for a given k_1H.

(4) LAUNCHING EFFICIENCY

The launching efficiency η can easily be evaluated if P_R/P_S is known. We have

$$\eta = \frac{P_S}{P_S + P_R} = \frac{1}{1 + P_R/P_S} . \quad . \quad . \quad (20)$$

In Fig. 3, η is plotted as a function of k_1H. It will be noticed that the curve for $k_1/k = 0$ ($kH \to \infty$) apparently sets an approximate lower limit to the launching efficiency.
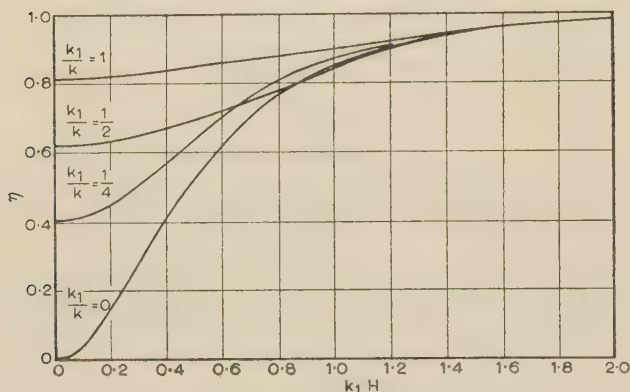
Fig. 3.—Launching-efficiency curves.

It may be of interest to mention that the series in eqn. (18) was in fact used for the computation, although for smaller values of k_1/k direct numerical integration of eqn. (17) would be quicker.

(5) CONCLUSIONS

Curves for the launching efficiency of a 'chopped surface-wave' aperture distribution have been obtained, and it is shown that, for all values of k_1/k, usefully high launching efficiencies are obtainable when the aperture height H is large enough to make $k_1H \sim 1$.

(6) REFERENCES

(1) CULLEN, A. L.: 'The Excitation of Plane Surface Waves', *Proceedings I.E.E.*, Monograph No. 93 R, February, 1954 (**101** Part IV, p. 225).
(2) RICH, G. J.: 'The Launching of a Plane Surface Wave', *ibid.*, Paper No. 1783 R, March, 1955 (**102** B, p. 237).
(3) Discussion on Reference 2, *ibid.*, 1955, **102** B, p. 824.

(7) APPENDIX

We need to evaluate the following integral:

$$I = \int_0^{\pi/2} \frac{\sin^2(kH\sin\theta)}{k_1^2 + k^2\sin^2\theta}d\theta = I' - I'' . \quad . \quad . \quad (21)$$

where

$$I' = \frac{1}{2}\int_0^{\pi/2} \frac{d\theta}{k_1^2 + k^2\sin^2\theta} = \frac{\pi}{4\beta_1 k_1} \quad . \quad . \quad (22)$$

$$I'' = \frac{1}{2}\int_0^{\pi/2} \frac{\cos(2kH\sin\theta)}{k_1^2 + k^2\sin^2\theta}d\theta \quad . \quad . \quad . \quad (23)$$

Now let $X = 2kH$, and note that

$$\cos(X\sin\theta) = J_0(X) + 2\sum_{m=1}^{\infty} J_{2m}(X)\cos 2m\theta . \quad (24)$$

Substituting eqn. (24) in eqn. (23), and carrying out the integration with respect to θ by using the result

$$\int_0^{\pi} \frac{\cos nx}{1 + a^2 - 2a\cos x}dx = \pi\frac{a^n}{1 - a^2}$$

we find, after some straightforward algebra, the result

$$I = \frac{\pi}{4\beta_1 k_1}\left[1 - J_0(x) - 2\sum_{m=1}^{\infty} J_{2m}(X)\left(\frac{\beta_1 - k_1}{k}\right)^{2m}\right] . \quad (25)$$

We next show that

$$\lim_{X\to\infty} I = \frac{\pi}{4\beta_1 k_1} \quad . \quad . \quad . \quad . \quad (26)$$

First note that $J_0(X) \sim X^{-\frac{1}{2}}$ for large X so $J_0(X) \to 0$ as $X \to \infty$. The same result holds for $J_{2m}(X)$ for any finite m in the series in eqn. (25), provided that $X \gg 2m$, but we must first consider X large but finite, and then carry out the summation to infinite value of m. It is not immediately obvious that letting $X \to \infty$ will make the infinite series vanish, since we know that

$$J_0(X) + 2\sum_{m=1}^{\infty} J_{2m}(X) = 1 \quad \text{for any } X.$$

The proof depends on the fact that $(\beta_1 - k_1)/k < 1$ for all values of k_1/k.

We first select a large value M for m such that

$$[(\beta_1 - k_1)/k]^{2M} \ll 1$$

The contribution to the summation from terms for which $m > M$ can then be made as small as desired by taking M sufficiently large. Provided that we then choose $X \gg M$, the previously used argument based on the asymptotic representation of $J_{2m}(X)$ can be employed for terms for which $m < M$, and the result of eqn. (26) follows.

An alternative argument making eqn. (26) plausible depends on the fact that eqn. (23) has an integral which is rapidly oscillatory, so that the integral would be expected to tend to zero as kH tended to infinity.

INTERMODULATION DISTORTION DUE TO FADING IN FREQUENCY-MODULATION FREQUENCY-DIVISION MULTIPLEX TRUNK RADIO SYSTEMS

By R. G. MEDHURST, B.Sc., and MURIEL HODGKINSON, B.Sc.

SUMMARY

Increased intermodulation distortion in frequency-modulation frequency-division multiplex trunk radio systems will be generated under fading conditions when the fade is due to the presence of one or more echoes. The distortion level is not simply related to the fading depth, but depends in a rather complicated way on the parameters of the echoes responsible for the fade. As a guide for future experimental work, theoretical investigation has been made of the nature of this dependence, both for single-echo and for multi-echo fading. In the latter case, a simple formula has been developed for the worst intermodulation distortion due to a system of echoes of given amplitudes and delay times. Numerical results are given for a particular $22 \cdot 5$ dB fade whose echo structure was derived experimentally.[2]

It is concluded that nothing can be deduced from measurements of fading depth alone about the magnitude of intermodulation distortion associated with fades generated by cancelling echoes. Measurements are required either of the parameters of the echoes or, directly, of the distortion.

LIST OF SYMBOLS

ω_M = Frequency modulation, rad/s.

ω_D = 'Peak' frequency deviation (taken as 11 dB above r.m.s.), rad/s.

ω_m = A base-band frequency, rad/s.

$\hat{\omega}_m$ = Maximum base-band frequency, rad/s.

$k = \omega_m/\hat{\omega}_m$.

ω_c = Carrier frequency, rad/s.

V_D/V_U = Distortion/signal voltage ratio in any one channel.

V_{D2}/V_U = Second-order distortion/signal ratio in any one channel.

V_{D3}/V_U = Third-order distortion/signal ratio in any one channel.

ϕ_D = Phase distortion due to echoes, rad.

$1/d$ = Instantaneous fading depth (i.e. the ratio of the undisturbed signal amplitude to that under fading conditions), voltage ratio.

For a single-echo fade:

r = Relative echo amplitude, voltage ratio.

τ = Echo delay, sec.

For a multiple-echo fade:

r_n = Relative amplitude of the nth echo, voltage ratio.

τ_n = Delay of the nth echo, sec.

(1) INTRODUCTION

In trunk radio systems two undesirable effects are associated with fading. First, there will be an increase in the ratio of r.f. noise to r.f. signal strength, with consequent increase in the relative noise level in any one telephone channel after demodulation. This effect can be calculated with good accuracy in a simple way: when the r.f. noise spectral distribution is flat, the well-known triangular distribution of noise appears after demodu-

lation.[1] The noise level per channel is simply related to the instantaneous fading depth.

The second effect relates to intermodulation distortion of the base band. An increase in intermodulation distortion will be associated with a fade if the fade is due to destructive interference with the main signal by one or more echoes. Such a fade is characterized by a fading depth which is dependent on frequency. It has been shown[2] for a particular overland path that all the deep fades studied were made up of two components, one frequency-dependent and the other frequency-independent. The latter is presumably due to a temporary increase in attenuation or to a bending of the main beam. It is to be expected that intermodulation distortion will be associated only with the frequency-dependent component of the fade.

Intermodulation distortion associated with frequency-selective fades is not related to the instantaneous fading depth in such a simple way as distortion due to r.f. noise. For example, in the simple case of a fade due to one short-delay echo, the distortion/signal voltage ratio in one channel, besides being proportional to the square of the delay time, depends on a factor (involving the echo amplitude and phase) which, for a given instantaneous fading depth, can vary from nearly zero to a finite maximum. When the fade is due to a number of echoes, the situation is correspondingly more complicated, the distortion depending quite critically on the amplitudes, delays and phase angles of the echoes.

There is a further important difference between the two deleterious effects of fading. The increase in relative r.f. noise level can be mitigated by increase in transmitter power or aerial gain, or by improvement in receiver noise factor. None of these measures will affect the increased intermodulation distortion, reduction of which will require an instantaneous change of path or frequency.

Although there are considerable (not very concordant) data on the frequencies and depths of fades over various paths at ultra-high frequency,[2-9] little has been published on the echo parameters associated with fading.[2] Nor, it appears, are any measurements available on the intermodulation distortion associated with fades. Consequently, it is of interest to see how far it is possible to obtain a theoretical indication of the orders of magnitude involved.

The only previously available theoretical work on intermodulation distortion due to fading appears to be that in Reference 10, where an expression is obtained for the worst distortion, at a given instantaneous fading depth, due to a single echo. In the present paper, the mechanism of distortion generation during single-echo fades is examined in considerably more detail and, incidentally, an expression for the worst distortion somewhat different from that of Reference 10 is derived. Since, at least on overland paths, fades tend to be due to a number of echoes of substantially different delays rather than to a single echo (Reference 2), an approximate expression is obtained for the worst distortion in the multiple-echo case. This is compared with the exact distortion level worked out in a specific case (considered in Reference 2) of a fade generated by six echoes.

(2) SINGLE-ECHO FADES

(2.1) Dependence of Distortion on Echo Phase

A single echo capable of causing a deep fade must have an amplitude close to that of the main signal and a phase difference from the main signal close to π. Also, such large echoes will be expected to have short delays. When the delay is sufficiently small to make both $\omega_D\tau/(1 - r)$ and $\hat{\omega}_m/(1 - r)$ very much less than unity (as will usually be the case), the distortion/signal voltage ratio, V_D/V_U, in any one channel, will be

$$\frac{V_D}{V_U} = \sqrt{\left[\left(\frac{V_{D2}}{V_U}\right)^2 + \left(\frac{V_{D3}}{V_U}\right)^2\right]} \quad \cdots \quad (1)$$

where V_{D2}/V_U and V_{D3}/V_U are given by[11]

$$\frac{V_{D2}}{V_U} = \frac{0\cdot2\tau^2 r(1 - r^2)\omega_D\omega_m\sqrt{(1 - 0\cdot5k)}\sin\omega_c\tau}{(1 + r^2 + 2r\cos\omega_c\tau)^2} \quad \cdots \quad (2)$$

$$\frac{V_{D3}}{V_U} = \frac{0\cdot028\tau^3 r(1 - r^2)(\cos\omega_c\tau + 4r - 2r\cos^2\omega_c\tau + r^2\cos\omega_c\tau)\omega_D^2\omega_m\sqrt{(1 - 0\cdot333k^2)}}{(1 + r^2 + 2r\cos\omega_c\tau)^3} \quad \cdots \quad (3)$$

When r is close to unity, as required for deep fading due to a single echo, and τ is of the order of magnitude usually encountered with large echoes, the distortion given by eqn. (1) will be negligible over most of the available range of echo phase angle, but two very sharp peaks of distortion, close to and on either side of the anti-phase condition, put in an appearance. A numerical example is shown in Fig. 1, the following parameters being used:

$$r = 0\cdot9$$
$$\tau = 0\cdot3 \text{ millimicrosec}$$
$$\omega_D = 4 \times 2\pi \times 10^6 \text{ rad}$$
$$\hat{\omega}_m = 2\cdot54 \times 2\pi \times 10^6 \text{ rad/s}$$
$$k = 1$$

The last three quantities refer to the top channel of a 600-channel system. The delay time corresponds to reflection (without change of phase) from the boundary of the first Fresnel zone when the carrier frequency is 1 667 Mc/s, and the echo amplitude is such as to generate a 20 dB fade at this frequency.

(2.2) Distortion at a Given Instantaneous Fading Depth

Taking the vector sum of the main signal and the echo, we have

$$d^2 = 1 + r^2 + 2r\cos\omega_c\tau \quad \cdots \quad (4)$$

For sufficiently short delay times V_{D3}/V_U will be negligible compared with V_{D2}/V_U. Then, from eqns. (1) and (2),

$$\frac{1}{\tau^2}\frac{V_D}{V_U} = \frac{0\cdot2\omega_D\omega_m\sqrt{(1 - 0\cdot5k)}}{d^4}r(1 - r^2)\sin\omega_c\tau \quad (5)$$

Making use of eqn. (4), this becomes

$$\frac{1}{\tau^2}\frac{V_D}{V_U} = \frac{0\cdot2\omega_D\omega_m\sqrt{(1 - 0\cdot5k)}}{d^4}\sin\omega_c\tau$$

$$\times\left[-\cos\omega_c\tau \pm \sqrt{(d^2 - \sin^2\omega_c\tau)}\right]$$

$$\times\left[2\sin^2\omega_c\tau - d^2 \mp 2\cos\omega_c\tau\sqrt{(d^2 - \sin^2\omega_c\tau)}\right] \quad (6)$$

The reason for the ambiguity of sign is that for certain ranges of phase angle the same fading depth can be generated by two echo amplitudes. When this is true, it is evident that two distortion levels can be associated with a given fading depth and echo phase angle. It can be readily shown that a given instantaneous fading depth, $1/d$, can only be generated by a single

echo if the echo phase angle lies within a region centred on π of width $2\{\pi - \arccos[-\sqrt{(1 - d^2)}]\}$. Within the narrower region, centred on π, of width $2\{\pi - \arccos[-(1 - \frac{1}{2}d^2)]\}$, only one set of signs in eqn. (6) is appropriate, while in the zone lying within the broader region and outside the narrower region both signs are permissible. This ambiguity of distortion level near the edges of the allowed range of phase angle, can be seen, for example, in Fig. 2.

Figs. 2, 3 and 4 show the variation with echo phase angle of $\tau^{-2}V_D/V_U$ for instantaneous fading depths of 10, 20 and 30 dB in the case of a 600-channel system subjected to fading generated by a single echo. These graphs show clearly that no unique distortion level can be associated with a given fading depth, even when the echo delay is known.
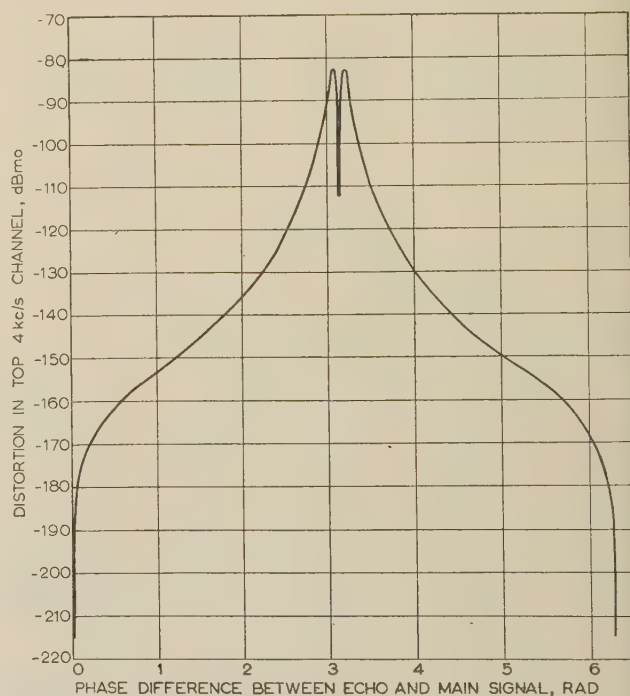
Fig. 1.—Variation with echo phase of the distortion due to a single echo.

Number of channels	..	600
Echo amplitude	..	0·9
Echo delay time	..	0·3 millimicrosec

(2.3) Maximum Distortion at a Given Instantaneous Fading Depth

A fairly simple expression can be derived for the maximum distortion possible at a given instantaneous fading depth when the fading is due to a single echo. It is (see Section 5.1)

$$\frac{1}{\tau^2}\left(\frac{V_D}{V_U}\right)_{max} = 0\cdot2\omega_D\omega_m\sqrt{(1 - 0\cdot5k)}$$

$$\frac{[-3d + \sqrt{(d^2 + 32)}][32 - 2d^2 - 2d\sqrt{(d^2 + 32)}]^{1/2}}{32d^2} \quad \cdots \quad (7)$$

this maximum occurring when

$$r = \sqrt{\left[\frac{3d^2 + 4 - \sqrt{(d^4 + 32d^2)}}{4}\right]} \quad . \quad . \quad . \quad (8)$$

and

$$\cos \omega_c \tau = \frac{d^2 - 8 + \sqrt{(d^4 + 32d^2)}}{4\sqrt{[3d^2 + 4 - \sqrt{(d^4 + 32d^2)}]}} \quad . \quad . \quad (9)$$

When $d \ll 1$, eqn. (7) takes the approximate form

$$\frac{1}{\tau^2}\left(\frac{V_D}{V_U}\right)_{max} = \frac{0 \cdot 2}{d^2}\omega_D \omega_m \sqrt{(1 - 0 \cdot 5k)} \quad . \quad . \quad (10)$$

This approximation exaggerates the maximum distortion to the extent shown in Table 1.
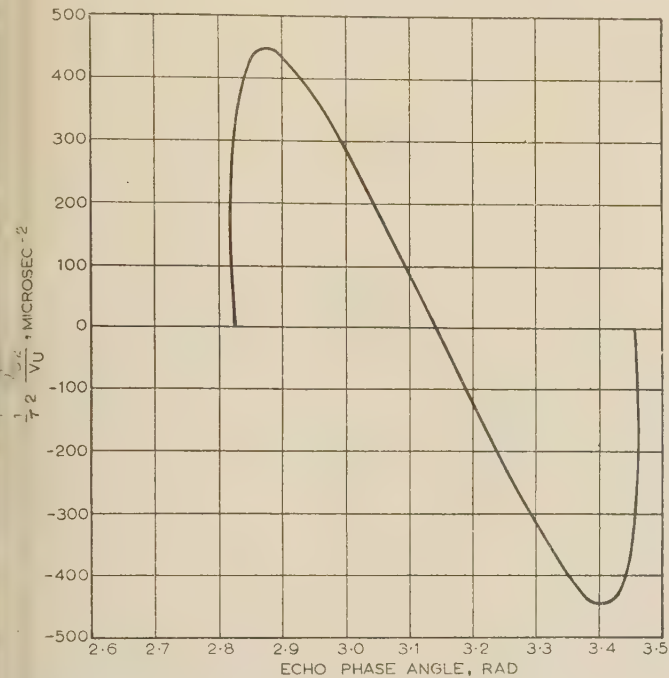
Fig. 2.—Variation with phase of second-order distortion for a
fading depth of 10 dB.

Single-echo fading: distortion in top channel of a 600-channel system.
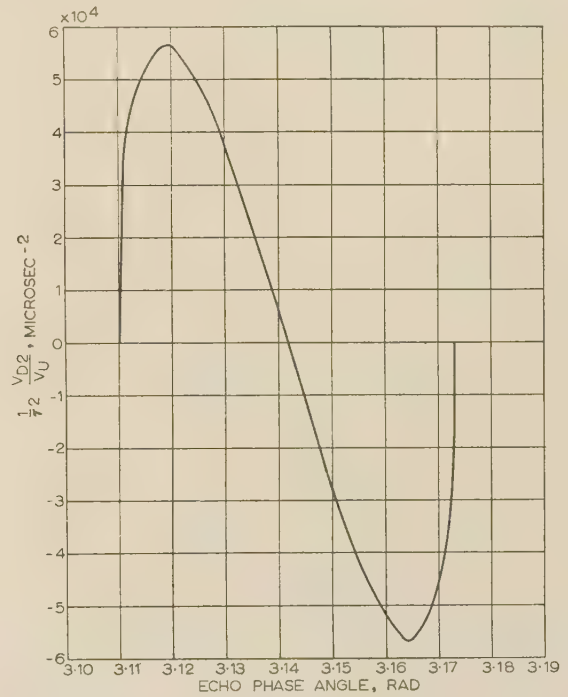
Fig. 4.—Variation with phase of second-order distortion for a fading
depth of 30 dB.

Single-echo fading: distortion in top channel of a 600-channel system.

Table 1

EXAGGERATION OF MAXIMUM DISTORTION DUE TO EQN. (10)

Instantaneous fading depth (d)	Excess of distortion given by approximate expression (10) over true distortion
dB	dB
10	2·1
20	0·6
30	0·2

Fig. 3.—Variation with phase of second-order distortion for a fading
depth of 20 dB.

Single-echo fading: distortion in top channel of a 600-channel system.

From eqn. (7), curves have been plotted (Fig. 5) showing the variation with echo delay of the worst distortion associated with instantaneous fading depths of 10, 20 and 30 dB, for the top channel of a 600-channel system. The boundaries of successive odd-number Fresnel zones are shown, for a carrier frequency of 1 800 Mc/s. While the curves have been drawn continuously, only the points of intersection with the odd-number Fresnel-zone boundaries will have physical significance, since reflections originating from regions lying between these boundaries will not generate either deep fades or appreciable distortion. However, since the locations of the boundaries are frequency-dependent, the solid curves are required when considering the variation of fading distortion with carrier frequency.

Fig. 5.—Variation with echo delay of maximum distortion associated with single-echo fading.

600-channel system.

(2.4) Expression by Albersheim and Schafer for Maximum Distortion

Albersheim and Schafer have given an expression for maximum distortion [expression (15) of Reference 10] which is of a form similar to eqn. (10) above but contains a smaller multiplying factor, their formula giving distortion levels nearly 4 dB lower than those computed from eqn. (10). The reason for this apparent discrepancy is that the two expressions do not represent quite the same physical quantity. Whereas eqn. (10) represents the worst distortion that can be associated with a given fading depth, for single-echo fading, Albersheim and Schafer considered the worst distortion generated by an echo of fixed amplitude and variable phase (the peak, in fact, of curves such as that shown in Fig. 1). Furthermore, the quantity $1/\delta$ in their formula, which replaces the $1/d$ in eqn. (10), is not the fading depth when the echo is phased for worst distortion, but the fading depth when the echo is in anti-phase.

(3) MULTIPLE-ECHO FADES

The existing evidence, based on one overland optical path,[2] suggests that fading due to echo interference on radio links of this kind generally involves a number of echoes of very varied delay times and amplitudes. It would not be expected that conclusions based on single-echo theory would be particularly relevant to this complicated situation. Fortunately, it is possible to derive a single approximate formula for the worst distortion that is capable of being generated by a number of echoes phased to produce a fade of given depth. This formula is a generalization of eqn. (10), and is of such a form as to show by inspection the way in which the contribution of each echo depends on its amplitude and delay.

(3.1) General Distortion Formula

It is supposed that all delay times are so short that only second-order distortion is significant. Then, it can be shown (Section 5.2)

that the distortion/signal voltage ratio in a channel of frequenc[y] ω_m is given by

$$\frac{V_D}{V_U} = 0\cdot4\omega_D\omega_m\sqrt{(1-0\cdot5k)}$$

$$\frac{[(BG-CF)(FG+BC)-(C^2+G^2)(AG-CE)]}{(C^2+G^2)^2} \quad . \quad (1[1])$$

where $A = -\frac{1}{2}\Sigma r\tau^2 \sin \omega_c\tau$
   $B = \Sigma r\tau \cos \omega_c\tau$
   $C = \Sigma r \sin \omega_c\tau$
   $E = -\frac{1}{2}\Sigma r\tau^2 \cos \omega_c\tau$
   $F = -\Sigma r\tau \sin \omega_c\tau$
   $G = 1 + \Sigma r \cos \omega_c\tau$

r and τ being the echo amplitudes and delays, and the summation extending over all the echoes.

(3.2) Approximate Formulae for Worst Distortion

The exact formula written down in the last Section is not o[f] great practical use, both because of its complexity and becaus[e] it requires detailed knowledge of the individual echo phases, a[s] well as of their amplitudes and delay times. Consequently, w[e] try to find a formula for the worst distortion at a given fadin[g] depth, analogous to eqn. (10). To do this (for the analytica[l] details see Section 5.2), consider a fixed instantaneous fadin[g] depth and suppose that all the echoes but one have fixed ampl[i]tudes and phases. Suppose also, for convenience, that th[e] variable echo has the smallest delay time, τ_1, say. We ca[n] now adjust the amplitude and phase of this variable echo (keep[ing] d fixed) to give maximum distortion, which will be a functio[n] of the amplitudes, delays and phase angles of the remainin[g] echoes. The final step is to allow these remaining phase angle[s] to vary until the maximum distortion just obtained has its larges[t] value. The expression so obtained is not a symmetrical func[tion] tion of all the echoes, since one echo has been subjected t[o] treatment different from that of the others. It is shown that i[n] order to generate a fade associated with maximum distortion, fo[r] a given fading depth, the sum of the echo amplitudes must b[e] close to unity. Inserting this condition, we arrive (Section 5.2[)] at the symmetrical formula

$$\left(\frac{V_D}{V_U}\right)_{max} = \frac{0\cdot2}{d^2}(r_1\tau_1 + r_2\tau_2 + \ldots)^2\omega_D\omega_m\sqrt{(1-0\cdot5k)} \quad . \quad (12)$$

This expression is valid only for small d, the approximation[s] being similar in nature to those applying to eqn. (10). If all th[e] τ's are made equal, and $\Sigma r_n = 1$, it will be noticed that th[e] expression reduces, as would be expected, to the form of eqn. (10[)]

(3.3) Numerical Example

In Reference 2, one specimen fade has been examined i[n] detail. It is concluded that the curve of fading depth agains[t] frequency (Fig. 4 of Reference 2) can be synthesized with goo[d] accuracy by addition to the main signal of the six echoes a[s] shown in Table 2. This combination of echoes produces [a] maximum fading depth of 22·5 dB, at about 4 005 Mc/s, givin[g] $d = 0\cdot075$. Also, $\Sigma r\tau = 1\cdot138 \times 10^{-9}$. Then, for the to[p] channel of a 600-channel system, an approximate estimate of th[e] worst distortion/signal value is found, from eqn. (12), to b[e] $-37\cdot8$ dB, corresponding to a distortion level in the top 4 kc/[s] channel of about -51 dBmo.*

To see how well this accords with the actual situation, values o[f] distortion in the neighbourhood of the frequency of deepest fad[e]

* 'dBmo' signifies decibels referred to a milliwatt measured at a specified point [o]n the telephony system: see, for example, Reference 12.

Table 2
SYSTEM OF ECHOES GENERATING A TYPICAL FADE*

Echo number	Echo amplitude (normalized to that of main signal) r	Echo delay τ	Phase shift from main signal	
			Half-wavelengths	Frequency
		mμs		Mc/s
1	0·45	0·122	1	4 090
2	0·26	0·370	3	4 050
3	0·10	2·86	23	4 033
4	0·115	3·14	25	3 981
5	0·025	3·9	31	3 972
6	0·02	12·1	97	3 993

* Measured by R. L. Kaylor (Bell Telephone Laboratories) in Iowa, U.S.A.
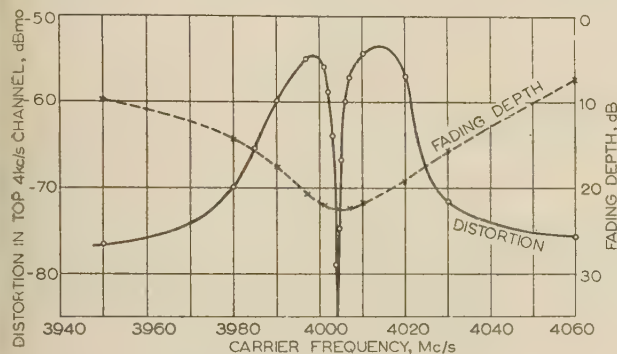
Fig. 6.—Distortion in a 600-channel system due to the set of six echoes described in Table 2.

have been evaluated from the exact formula (11), the resulting curve being shown in Fig. 6. It is seen that eqn. (12) gives a distortion level within about 3 dB of the true worst distortion. As in the single-echo case, substantial intermodulation distortion is generated over only part of the fading range: in particular, at the frequency of deepest fade the distortion falls steeply to a very low level.

In Reference 2, it was pointed out that omission of the two smaller echoes produces a substantial change in the shape of the fading curve. It is now also apparent, from expression (12), that these echoes make an important contribution to the distortion, in spite of their small amplitudes, since the significant parameter affecting distortion is the product of amplitude and delay, and the long delays compensate for the low amplitudes.

The significance of distortion levels such as those shown in Fig. 6, from the point of view of trunk radio system design, depends on the required system performance, the length of the system, and the percentage of time during which the distortion occurs. Not enough experimental data exist to clarify the latter aspect. During the series of measurements reported in Reference 2, fades of 22·5 dB or more occurred for only about 0·2% of the time. However, it would be quite unsafe to conclude, on the basis of an analysis of one not necessarily typical fade, that the worst distortion level (−54 dBmo) encountered in that case will occur for less than 0·2% of the time, even over the particular path under test. Regarding required system performance, some idea of order of magnitude is obtained by observing that a tentative C.C.I.R. recommendation,[13] based on C.C.I.F. requirements for cable transmission,[14] leads to the conclusion that the total intermodulation distortion per channel in a 180-mile system forming part of an international network should not exceed −57 dBmo for more than 99% of the time. Since the worst distortion generated during the fade considered here exceeds this permitted level by 3 dB, it is evident

that intermodulation distortion generated in fades may be of concern in system design, and further experimental work appears to be called for.

(4) REFERENCES

(1) CROSBY, M. G.: 'Frequency Modulation Noise Characteristics', *Proceedings of the Institute of Radio Engineers*, 1937, **25**, p. 472.

(2) KAYLOR, R. L.: 'A Statistical Study of Selective Fading of Super-High Frequency Radio Signals', *Bell System Technical Journal*, 1953, **32**, p. 1187.

(3) DURKEE, A. L.: 'Results of Microwave Propagation Tests on a 40 Mile Overland Path', *Proceedings of the Institute of Radio Engineers*, 1948, **36**, p. 197.

(4) CRAWFORD, A. B., and JAKES, W. C., Jr.: 'Selective Fading of Microwaves', *Bell System Technical Journal*, 1952, **31**, p. 68.

(5) BRYAN, L. A., and MILLAR, J. Z.: 'Notes on Propagation', *Convention Record of the Institute of Radio Engineers*, National Convention, 1953, Pt. 2, Antennas and Communications, p. 68.

(6) BOYD, J. E., and BROWN, F. B.: 'Propagation of 3, 10 and 23 cm Waves over a 50-Mile Optical Path', Georgia Institute of Technology, State Engineering Experimental Station, Atlanta, Georgia, Technical Report No. 1, Project No. 109–8, Contract No. W28–099–ac–175, 1950 (March).

(7) MATSUO, S., UGAI, S., KAPITA, K., IKEGAMI, F., and KONO, Y.: 'Microwave Fading', *Reports of the Electrical Communication Laboratories*, Nippon Telegraph and Telephone Public Corporation, Japan, 1953, **1**, p. 38.

(8) CHAVANCE, P., BOITHIAS, L., and BLASSEL, P.: 'Étude de propagation d'ondes centimétriques dans la région Méditerranéenne', *Annales des Telecommunications*, 1954, **9**, p. 158.

(9) MATHWICH, H. R., NUTTALL, E. D., PITMAN, J. E., and RANDOLPH, A. M.: 'Propagation Test on Microwave Communication Systems', *Electrical Engineering*, 1956, **75**, p. 1020.

(10) ALBERSHEIM, W. J., and SCHAFER, J. P.: 'Echo Distortion in the F.M. Transmission of Frequency-Division-Multiplex', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 316.

(11) MEDHURST, R. G., and SMALL, G. F.: 'An Extended Analysis of Echo Distortion in the F.M. Transmission of Frequency-Division Multiplex', *Proceedings I.E.E.*, Paper No. 2006 R, March, 1956 (**103 B**, p. 190).

(12) FISHER, W. C.: 'Transmission and Power Levels in Multiplex Equipment', *Radio-Electronic Engineering*, 1954, **23**, p. 10.

(13) C.C.I.R. Documents of the VIIth Plenary Assembly, London, 1953, I, Recommendation No. 40, p. 35.

(14) Comité Consultatif International Téléphonique (C.C.I.F.), XVIᵉ Assemblée Plénière, Firenze, 1951, III *bis*; Transmission sur les lignes maintenance, p. 121.

(5) APPENDICES

(5.1) Maximum Intermodulation Distortion at a Given Instantaneous Fading Depth for Single-Echo Fading

Eqn. (7) is obtained most straightforwardly by eliminating $\omega_c\tau$ between eqns. (4) and (5), and maximizing with respect to r. From eqns. (4) and (5) we have

$$\frac{1}{\tau^2}\frac{V_D}{V_U} = \frac{0\cdot2\omega_D\omega_m\sqrt{(1-0\cdot5k)}}{2d^4}(1-r^2)[4r^2-(d^2-1-r^2)^2]^{1/2}$$

$$\cdots\cdots (13)$$

Differentiating with respect to r^2 and equating to zero, we find stationary values where

$$2r^4 - r^2(3d^2 + 4) + d^4 - d^2 + 2 = 0$$

giving
$$r^2 = \frac{3d^2 + 4 - \sqrt{(d^4 + 32d^2)}}{4} \quad . \quad . \quad . \quad (14)$$

(the minus sign being chosen to make $r^2 < 1$). Thus eqn. (8) follows. Substitution in eqn. (13) gives eqn. (7), and elimination of r between eqns. (4) and (14) gives eqn. (9).

(5.2) Approximate Expression for the Worst Intermodulation Distortion in the Multiple-Echo Case

When all the delay times are sufficiently small, the phase error in the multiple-echo case (by a straightforward extension of Appendix 1 of Reference 10) is given by

$$\phi_D = - \arctan \left[\frac{\Sigma r \sin (\omega_c \tau + \omega_M \tau)}{1 + \Sigma r \cos (\omega_c \tau + \omega_M \tau)} \right]$$

Thus, approximately,

$$\phi_D =$$
$$- \tan \arc \left\{ \frac{\Sigma[r(1 - \tfrac{1}{2}\omega_M^2 \tau^2) \sin \omega_c \tau + r \omega_M \tau \cos \omega_c \tau]}{1 + \Sigma[r(1 - \tfrac{1}{2}\omega_M^2 \tau^2) \cos \omega_c \tau - r \omega_M \tau \sin \omega_c \tau]} \right\}$$
$$. \quad . \quad . \quad (15)$$

$$= - \arctan \left(\frac{A\omega_M^2 + B\omega_M + C}{E\omega_M^2 + F\omega_M + G} \right)$$

where A, B, C, E, F, G have the significance defined in Section 3.1. When this expression is expanded in powers of ω_M, the coefficient of ω_M^2 is given by

$$\frac{1}{2}\frac{d^2\phi_D}{d\omega_M^2}\bigg|_{(\omega_M = 0)} = \frac{1}{(C^2 + G^2)^2}$$

$$[(BG - CF)(FG + BC) - (C^2 + G^2)(AG - CE)] \quad . \quad (16)$$

From this, eqn. (11) can be immediately deduced.[10]

Now, put

$$\left. \begin{array}{l} 1 + \Sigma r \cos \omega_c \tau = d_c \\ \Sigma r \sin \omega_c \tau = d_s \end{array} \right\} \quad . \quad . \quad . \quad (17)$$

and

so that the reciprocal of the instantaneous fading depth is given by

$$d^2 = d_c^2 + d_s^2$$

We can eliminate the r and the $(\omega_c \tau)$ associated with one echo from eqn. (15) by the use of eqns. (17). Calling the eliminated quantities r_1 and $\omega_c \tau_1$, eqn. (15) becomes

$$\phi_D = - \arctan \left[\frac{d_s - \tfrac{1}{2}\omega_M^2(\tau_1^2 d_s + J) + \omega_M(\tau_1 d_c + K)}{d_c - \tfrac{1}{2}\omega_M^2(\tau_1^2 d_c + L) - \omega_M(\tau_1 d_s + M)} \right]$$

where

$$J = r_2(\tau_2^2 - \tau_1^2) \sin \omega_c \tau_2 + r_3(\tau_3^2 - \tau_1^2) \sin \omega_c \tau_3 + \ldots$$
$$K = - \tau_1 + r_2(\tau_2 - \tau_1) \cos \omega_c \tau_2 + r_3(\tau_3 - \tau_1) \cos \omega_c \tau_3 + \ldots$$
$$L = - \tau_1^2 + r_2(\tau_2^2 - \tau_1^2) \cos \omega_c \tau_2 + r_3(\tau_3^2 - \tau_1^2) \cos \omega_c \tau_3 + \ldots$$
$$M = r_2(\tau_2 - \tau_1) \sin \omega_c \tau_2 + r_3(\tau_3 - \tau_1) \sin \omega_c \tau_3 + \ldots$$

Then, in the previous nomenclature,

$$A = - \tfrac{1}{2}(\tau_1^2 d_s + J)$$
$$B = \tau_1 d_c + K$$
$$C = d_s$$
$$E = - \tfrac{1}{2}(\tau_1^2 d_c + L)$$
$$F = - (\tau_1 d_s + M)$$
$$G = d_c$$

The term in square brackets in eqn. (16) becomes

$$(BG - CF)(FG + BC) - (C^2 + G^2)(AG - CE)$$
$$= [d_c(\tau_1 d_c + K) + d_s(\tau_1 d_s + M)]$$
$$\quad [- d_c(\tau_1 d_s + M) + d_s(\tau_1 d_c + K)]$$
$$\quad - (d_c^2 + d_s^2)[- \tfrac{1}{2}d_c(\tau_1^2 d_s + J) + \tfrac{1}{2}d_s(\tau_1^2 d_c + L)]$$
$$= [\tau_1(d_c^2 + d_s^2) + d_c K + d_s M](- d_c M + d_s K)$$
$$\quad + \tfrac{1}{2}(d_c^2 + d_s^2)(d_c J - d_s L) \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

We now assume that the d's are sufficiently small to allow suppression of the third-order terms [this is the same order o approximation as that used in deriving eqn. (10)]. Then expression (18) reduces to

$$(d_c K + d_s M)(- d_c M + d_s K) = (d_s^2 - d_c^2)KM + d_s d_c(K^2 - M^2)$$

Put
$$d_s = d \sin \theta$$
$$d_c = d \cos \theta$$

so that expression (18) becomes

$$d^2[\tfrac{1}{2}(K^2 - M^2) \sin 2\theta - KM \cos 2\theta] \quad . \quad . \quad (19)$$

Now, if all the echoes except the first have fixed amplitudes and phases, the problem of adjusting the amplitude and phase o the first echo for maximum distortion is reduced to that of making the proper choice of θ in order to maximize expression (19). This maximum value is immediately seen to be

$$d^2[\tfrac{1}{4}(K^2 - M^2)^2 + (KM)^2]^{\frac{1}{2}}$$
$$= \frac{d^2}{2}(K^2 + M^2)$$

We can now, finally, choose $\omega_c \tau_2$, $\omega_c \tau_3$, etc., to maximize $K^2 + M^2$. To fix our ideas, suppose that τ_1 is the least delay time. Then, $K^2 + M^2$ is the square of the amplitude of the resultant of a number of co-planar vectors of amplitudes τ_1, $r_2(\tau_2 - \tau_1)$, $r_3(\tau_3 - \tau_1)$, etc., and phase angles π, $\omega_c \tau_2$, $\omega_c \tau_3$ etc. The maximum value of this resultant amplitude is evidently obtained when

$$\omega_c \tau_2 = \omega_c \tau_3 = \ldots = \pi \quad . \quad . \quad . \quad (20)$$

so that

$$(K^2 + M^2)_{max} = [\tau_1 + r_2(\tau_2 - \tau_1) + r_3(\tau_3 - \tau_1) + \ldots]^2$$

Thus, the worst distortion/signal ratio takes the form

$$\left(\frac{V_D}{V_U} \right)_{max} = \frac{0 \cdot 2}{d^2}[\tau_1 + r_2(\tau_2 - \tau_1) + r_3(\tau_3 - \tau_1) + \ldots]^2$$
$$\omega_D \omega_m \sqrt{(1 - 0 \cdot 5k)} \quad . \quad (21)$$

According to condition (20), all the echoes except the first are required to be in phase opposition to the main signal. Thus, for a deep fade to exist, the first echo, if it is of substantial amplitude, must also be nearly in phase opposition (since there are no other available echoes having phases appropriate for cancellation of any considerable quadrature component of the first). Hence, for a deep fade producing maximum distortion, we have the approximate relation

$$r_1 + r_2 + \ldots \simeq 1$$

Inserting this condition into eqn. (21), we arrive at the symmetrical form

$$\left(\frac{V_D}{V_U} \right)_{max} \simeq \frac{0 \cdot 2}{d^2}(r_1 \tau_1 + r_2 \tau_2 + \ldots)^2 \omega_D \omega_m \sqrt{(1 - 0 \cdot 5k)}$$

which is expression (12).

A VARIABLE-VELOCITY SCANNER FOR MAGNETIC DEFLECTION OF A SCANNING SPOT

By Professor M. P. BEDDOES, B.Sc., Associate Member.

SUMMARY

Designed initially for use with a method for television band compression by variable-velocity scanning, the performance limitations of a magnetic scanner with negative feedback from a novel source are presented: also, some of the many applications of such a scanner are briefly reviewed.

The justification of the presentation relies mainly on the mathematical analysis: however, measurements of current waveforms confirm that the influence of non-linearities and other factors which have been explicitly ignored in the derivation of this analysis is negligible in practice.

Consideration is given to the practical application of the results, and it is shown that satisfactory operation can easily be obtained for repetitive scanning rates up to 10^4 per second with a maximum displacement error (during a change of velocity) of no more than $0 \cdot 2\%$.

LIST OF PRINCIPAL SYMBOLS

$V_{in}(t)$ = Signal voltage applied to amplifier.

$V_{fb}(t)$ = Feedback voltage.

$V_a(t) = V_{in}(t) - V_{fb}(t)$ = One input to the amplifier.

β = Feedback transfer ratio, voltage/current.

M = Feedback factor = $A_3\beta$.

m, A_1, A_2, A_3 = Gains of amplifiers.

$i(t)$ = Flux-producing current in deflecting coil reflecting at the amplifier output.

$i_a(t)$ = Driving current from the amplifier.

$\bar{i}$ = Laplace transform of $i(t)$.

$\bar{i}_a$ = Laplace transform of $i_a(t)$.

$\omega_0 = 2\pi f_0$ = Resonant angular frequency of deflecting coils and output transformer.

Ω_0 = Effective resonant angular frequency of deflecting coils when connected to the amplifier.

Q_s = Q-factor of the deflecting coil in circuit without external damping (Fig. 3).

$K_1 = dV_{in}(t)/dt, t < 0$
$K_2 = dV_{in}(t)/dt, t > 0$ } Fig. 4.

$U_1 = di/dt, t < 0$.

U_2 = Steady-state slope of $i(t)$, di/dt, for $t > 0$.

$n = U_1/U_2$.

T_s = Line period.

I_{pp} = Peak-to-peak deflecting current from the amplifier.

V_s = Peak voltage from the amplifier (series damping).

V_p = Peak voltage from the amplifier (shunt damping).

L, C, r = Deflecting-coil parameters (Fig. 3) reflected to the output of the amplifier.

R = External damping resistance.

I_{max} = Maximum error of $i(t)$ at a change of state.

$d = I_{pp}/I_{max}$ = Reciprocal of the maximum allowable error of deflecting current.

α_I = Fractional increase of output current when the amount of feedback is increased from zero to M.

α_ω = Fractional increase of the effective bandwidth when negative feedback is applied.

(1) INTRODUCTION

Variable-velocity scanning was used before the word 'television' was invented: for example, in 1909, Gilbert Sellers[1] patented a mechanical variable-velocity scanner for the transmission of facsimile messages by telegraphy. Bedford and Puckle[2] were the first to use a variable-velocity scanner in practical television. They proposed a system to modulate the brightness of the raster on a cathode-ray tube. According to their proposals, the local scanning velocity was varied inversely as the desired local brightness. Thus, in order to transmit a picture, information of the local scanning velocity only was required, because the current of the receiver scanning beam was held constant. The Bedford–Puckle variable-velocity scanner was essentially simple. The scanning-beam deflecting plate was connected to the junction of a pentode anode and a capacitor; the capacitor was discharged by the anode current of the pentode. In the absence of any incoming signal, the pentode grid was held at a constant potential. Thus, the voltage across the capacitor increased uniformly with time, and the spot was deflected at a uniform velocity across the cathode-ray-tube screen. If, at a particular point, the local scanning velocity was to be decreased (producing increased local brightness), the control grid was lowered in potential, and conversely. The Bedford–Puckle variable-velocity scanner, as it was described, was suitable only for driving cathode-ray tubes with electrostatic deflection plates: it could not be employed for magnetic deflection of the scanning spot.

Cherry and Gouriet[3] (and independently, Bell[4]) proposed a system for bandwidth compression of picture signals using variable-velocity scanners at transmitter and receiver. When scanning sections of the picture containing fine detail,* both transmitter and receiver scanning spots moved slowly; when scanning sections of the picture containing little or no detail, both spots moved rapidly. It was claimed that such a system should reduce the signal bandwidth below that normally required for a constant-velocity scanning system. The Cherry–Gouriet system was essentially a running coding system for reducing the redundancy between adjacent picture points. A somewhat similar variable-velocity system was proposed by Schroter.[6] According to his suggestions, bandwidth compression was achieved by reducing the frame information (frame-to-frame redundancy).

For variable-velocity operation in general[2–6] the scanning velocity is varied within a range of the same sign (i.e. the spot does not locally retrace its path). For certain applications,[3,4,6] the highest scanning velocity is approximately that used for

---

Correspondence on Monographs is invited for consideration with a view to publication.

Prof. Beddoes is Assistant Professor in the Department of Electrical Engineering, University of British Columbia, and was formerly in the Department of Electrical Engineering, Imperial College of Science and Technology, University of London.

* For definitions of 'picture detail' see References 3 and 7

constant-velocity scanning, and departures from this value are to lower values of the same sign. For these applications therefore, instead of increasing the deflecting-coil voltage, variable-velocity operation reduces it for certain periods of time.

In order to change the scanning rate with a magnetic scanner, the voltage across the deflecting coil must alter to a new value: the unavoidable stray capacitance of the coil opposes the rate at which its terminal voltage changes. Thus, a large driving pulse of current is required in order to effect a rapid change of scanning velocity: this pulse may materially increase the range of current required from the circuit driving the coil.

The purpose of the paper is to outline the main design features of a variable-velocity magnetic scanner. It will be shown that the design, in one of its phases, is practical and may be achieved without special components; in particular, without excessive insulation of the turns of the deflecting coil. Because of the satisfactory results of calculation, it is thought that this scanner may be used for any of the variable-velocity instruments briefly reviewed above.

(2) APPARATUS

The variable-velocity scanner consists of two parts: a driver sweep generator, and a driven amplifier which operates into the deflecting coil.

(2.1) Driver Sweep Generator

The sweep generator is similar in principle to the Miller valve, V_2, of the Rao–Sankava Subramanyan expanded time-base.[5] In
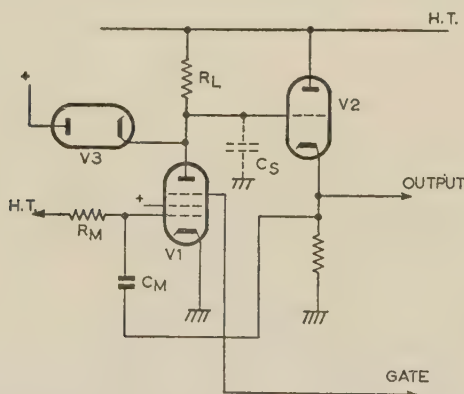
Fig. 1.—Circuit diagram for the sweep generator.

addition, however, the present generator uses a cathode-follower (Fig. 1), principally in order to limit the flyback time;[8] also, a diode is employed to limit the extent of the sweep stroke.

The scan speed is required to be locally varied by some given controlling voltage (i.e. the slope of the sweep stroke is to be varied by this voltage). For this operation, therefore, the contr[ol] voltage is applied to the grid resistance of V_1.

(2.2) The Main Amplifier

The amplifier which supplies the deflecting coil is itself drive[n] by the sweep generator. Before describing this amplifier, [an] outline will be given of the voltage/current relationships whic[h] exist across a deflecting coil.

For a deflecting coil of high Q-factor, a constant potenti[al] applied externally to its terminals produces a constant increa[se] of current with time: if the external potential is altered, th[e] scanning rate alters, and it can do so as rapidly as th[e] potential across the coil is altered. The sole hindrance t[o] rapid changes in coil potential (and hence to rapid chang[es] in scanning rate) is caused by the stray capacitance C_c: th[is] is the sum of the coil stray capacitance and that of th[e] circuit driving the coil. For a given value of C_c, the use of [a] highly permeable deflecting-coil yoke greatly increases the defle[c]tion sensitivity above that of the same coil in air. Conversel[y] for a given deflection sensitivity, the deleterious effects of C_c ca[n] be greatly reduced by the use of a highly permeable yoke. Th[e] undesired effect of C_c may be further reduced by decreasing th[e] deflecting-coil inductance L_d, to the point at which the inductanc[e] L_l of the leads to the coil becomes an appreciable fraction of L further reduction of L_d leads to inefficient operation, as a[n] appreciable fraction of deflecting power is then used to build u[p] the magnetic field in L_l.

The main deflecting coil (Fig. 2) is driven by the output ampl[i]fier A_3, which is connected to deliver constant current to the co[il] for a given input voltage V_a. The flux of the main coil links [a] monitor coil, hence the voltage v which appears across the latt[er] is proportional to the rate of change of useful flux-producir[ng] current i in the main coil: $v = K di/dt$, where K is a constan[t] The voltage v is fed to a Miller integrator, the output, V, of whic[h] is proportional to i. V is amplified and is used as negativ[e] feedback to the output amplifier A_3.

The monitor and the main deflecting coils are wound side b[y] side. The monitor coil has rather less than a tenth the numb[er] of turns of the main deflecting coil: hence, the self-resonanc[e] of the monitor coil is much higher than that of the mai[n] deflecting coil (e.g. in the present design the main deflectir[g] coil resonates with its stray capacitance at 400 kc/s, while th[e] monitor coil resonates at $1 \cdot 5$ Mc/s). Thus, the flux is faithful[ly] monitored past the self-resonance f_0 of the main coil.

(3) ANALYSIS OF THE EFFECT OF FEEDBACK ON TH[E] TRANSIENT PERFORMANCE OF THE AMPLIFIE[R] DRIVING THE DEFLECTING COIL

The claims made here for the variable-velocity magnetic scanne[r] are mainly justified by the mathematical analysis of its transier[t]
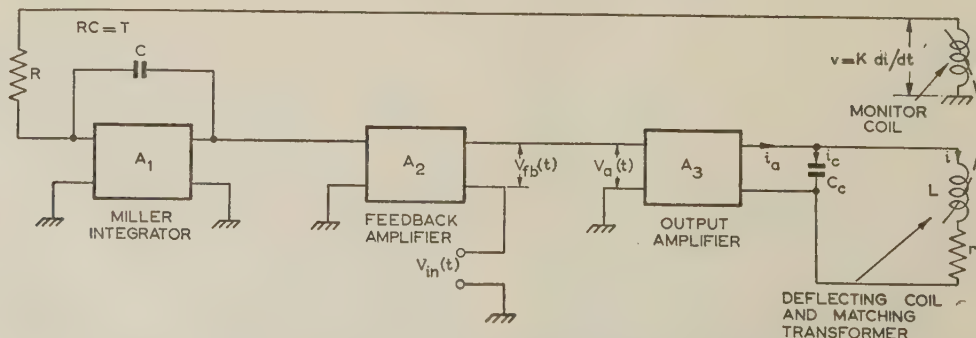
Fig. 2.—Simplified diagram of the main amplifier driving the deflecting coil.

performance. Standard Laplace transform methods[9] are employed to arrive at the solution; but, for the sake of conciseness, many intermediate steps are omitted. Where possible, the mathematical analysis is supplemented by oscillograms of current waveforms.

Six assumptions are made:

(a) All component values are independent of current loading.

(b) All electronic circuits are linear. For the output valve driving the deflecting coil, this assumption will be justified following the calculations of the output loading conditions (Section 4).

(c) All circuits (including the monitor loop) have bandwidths which considerably exceed that of the main deflecting coil, which is limited by the coil stray capacitance C_c.

(d) The current supplied to the deflecting coil is independent of the coil impedance.

(e) The analysis is concerned with the performance of the amplifier when the scanning velocity is suddenly changed. It is assumed that at any such instant, transients due to previous changes of velocity have decayed sufficiently for their effects to be neglected.

(f) It is assumed that the primary-to-secondary coupling of the output matching transformer is extremely close. Thus, for the purpose of analysis, the circuit for the transformer driving the deflecting coil may be replaced by a parallel-tuned circuit of low Q-factor (Fig. 3).
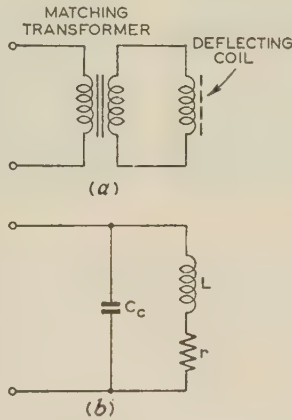
Fig. 3.—Series-damped deflecting coil and output transformer.
(a) Exact circuit diagram.
(b) Simplified equivalent diagram.

Two cases are analysed: (a) for series damping (Section 3.1); here, damping is considered to be achieved entirely by a resistance in series with the inductance of the deflecting coil; and (b) for shunt damping (Section 3.2); a resistor placed in shunt with the deflecting coil provides most of the required damping.

(3.1) Transient Analysis for Series Damping

A monitor coil links the flux of the main deflecting coil; the voltage from this monitor coil is integrated by a Miller integrator, and the resulting output is applied as negative feedback to the output amplifier A_3 (Fig. 2). The voltage $V_a(t)$ at the input of A_3 is thus given by

$$V_a(t) = V_{in}(t) - i(t)\beta \quad . \quad . \quad . \quad . \quad (1)$$

and the driving current $i_a(t) = -A_3 V_a(t)$. Also, for a high-gain Miller valve, the feedback ratio (dimensions of resistance) $\beta = KA_2/T$, where T is the time-constant of the Miller valve.[10] Thus, eqn. (1) may be rewritten as follows:

$$V_{in}(t) = i(t)\beta - \frac{i_a(t)}{A_3} \quad . \quad . \quad . \quad . \quad (2)$$

Now, the ratio of the flux-producing current to the driving current is given by

$$\frac{\bar{i}}{\bar{i}_a} = \frac{1}{\left(\dfrac{p}{\omega_0}\right)^2 + \dfrac{1}{Q_s}\left(\dfrac{p}{\omega_0}\right) + 1}$$

where $Q_s = \omega_0 L/r$ and $\omega_0^2 = 1/LC_c$ (Fig. 3).

Without feedback (see Fig. 2), $\bar{i}_a = \overline{V_{in}}A_3$, and, under these conditions,

$$\bar{i} = \overline{V_{in}}\left[\frac{A_3}{\left(\dfrac{p}{\omega_0}\right)^2 + \dfrac{1}{Q_s}\left(\dfrac{p}{\omega_0}\right) + 1}\right]$$

With negative feedback, however, the right-hand bracketed term must be substituted for m in the general expression for feedback:

$$\bar{i} = \overline{V_{in}}\frac{m}{1-m\beta} = \frac{\overline{V_{in}}A_3}{1-M}\frac{1}{\left(\dfrac{p}{\Omega_0}\right)^2 + \dfrac{1}{Q_s}\dfrac{1}{\sqrt{(1-M)}}\dfrac{p}{\Omega_0} + 1} \quad . \quad (3)$$

where $\Omega_0 = \omega_0\sqrt{(1-M)}$ and the feedback factor[12] is $M (= A_3\beta)$: this is negative for negative feedback.

For critical damping,

$$\frac{1}{Q_s\sqrt{(1-M)}} = 2$$

or

$$r = 2\omega_0 L\sqrt{(1-M)} \quad . \quad . \quad . \quad . \quad (4)$$

and the transform of the flux-producing current [eqn. (3)] is

$$\bar{i} = \frac{\overline{V_{in}}A_3}{1-M}\frac{\Omega_0^2}{(p+\Omega_0)^2} \quad . \quad . \quad . \quad . \quad (5)$$

A particular stylized input[11] V_{in} will be considered: this voltage increases uniformly at a rate K_1 for time $t < 0$, and at K_2 for $t > 0$ (Fig. 4): but in order to assess the effect of increasing the negative feedback when the steady-state terms of i are held constant, the following substitution to the output current slopes U_1 and U_2 is made:

$$U_1 = -\frac{A_3 K_1}{1-M}$$

and

$$U_2 = -\frac{A_3 K_2}{1-M} \quad . \quad . \quad . \quad . \quad (6)$$

Suppose that $V_{in}(t)$ is a step-input of height K_2; then, from eqn. (5),

$$i(t) = U_2[1 - \varepsilon^{-\Omega_0 t}(1 + \Omega_0 t)]$$

If, however, the input V_{in} had been K_1 for $t < 0$, and then was suddenly changed to K_2 for $t > 0$, superposition clearly shows that

$$i(t) = (U_2 - U_1)[1 - \varepsilon^{-\Omega_0 t}(1 + \Omega_0 t)] + U_1 \quad . \quad (7)$$

The response to the input of Fig. 4 is obtained by simply integrating eqn. (7): thus,

$$i(t) = U_2 t + i_0 - \varepsilon^{-\Omega_0 t}\left(\frac{2}{\Omega_0} + t\right)(U_1 - U_2) + \frac{2}{\Omega_0}(U_1 - U_2) \quad . \quad (8)$$

$\text{steady-state terms} \quad \text{transient terms} \quad \text{error term}$

where i_0 is the current at $t = 0$.

(3.1.1) Effect of Negative Feedback.

The driving current $i_a(t)$ is given by the following formula:

$$i_a(t) = [D^2 LC + DC_c R + 1]i(t) \quad . \quad . \quad (9)$$
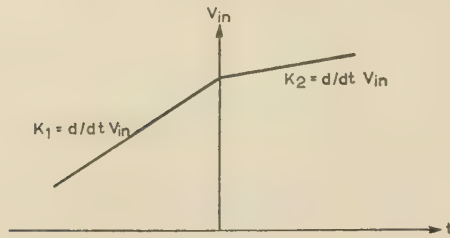
Fig. 4.—Input voltage drive to main amplifier.

where D is the operator d/dt, and L, C_c and R are shown in Fig. 3.

Eqns. (8) and (9) are combined to give the following expression for $i_a(t)$:

$$i_a(t) = U_2 t + i_0$$
steady-state terms

$$- (M \exp - \Omega_0 t)\left(\frac{2}{\Omega_0} + t\right)(U_1 - U_2)$$
transient terms

$$+ \frac{2M}{\Omega_0}(U_1 - U_2) \quad . \quad (10)$$
error term

The transient and error terms in eqns. (8) and (10), respectively, differ only by the multiplying factor M. This is negative for negative feedback. Thus, the transient and error pulse of driving current is M times the corresponding pulse in the deflecting coil; also, the sign of the driving pulse tends to diminish the magnitude of the pulse in $i(t)$.

(3.1.2) Maximum Error in Flux-Producing Current due to Overshoot.

There is a transient (overshoot) error in $i(t)$ each time the scanning velocity is changed. The maximum value of the error current, I_{max}, is obtained by differentiating the transient and error terms in eqn. (8) and equating to zero. By these means

$$I_{max} = -\frac{U_1 - U_2}{\omega_0\sqrt{(1 - M)}} \quad . \quad . \quad . \quad (11)$$

The magnitude of I_{max} is symmetrical with regard to U_1 and U_2. From this symmetry, it is apparent that the magnitude of the maximum error between scanning limits of the same sign is the same whether the scan rate is accelerated or decelerated.

(3.1.3) Deflecting-Coil Bandwidth for a given Tolerable Displacement Error.

The variable-velocity scanner may be considered to operate in the horizontal ('line') direction. Let the tolerable displacement error be $1/d$ of a line scan, let U_1 amperes per second correspond to the maximum scanning rate, and let T_s be the line period. The peak-to-peak scanning current $I_{pp} = U_1 T_s$: also, let the scanning velocity ratio be $n = U_1/U_2$ for $n > 1$. Then, from eqn. (11),

$$I_{max} = \frac{I_{pp}}{d} = \frac{U_1 T_s}{d} = \frac{U_1 - U_2}{\omega_0\sqrt{(1 - M)}}$$

whence the deflecting-coil bandwidth $f_0 = \omega_0/2\pi$ is given by

$$f_0 = \frac{d\left(1 - \frac{1}{n}\right)}{2\pi T_s\sqrt{(1 - M)}} \quad . \quad . \quad . \quad (12)$$

Eqn. (12) expresses the deflecting-coil bandwidth requirement for a given tolerable error. Fig. 7 shows the values of f_0 necessary for operating a variable-velocity scanner: the line frequency is

(a)   (d)

(b)   (e)

(c)   (f)
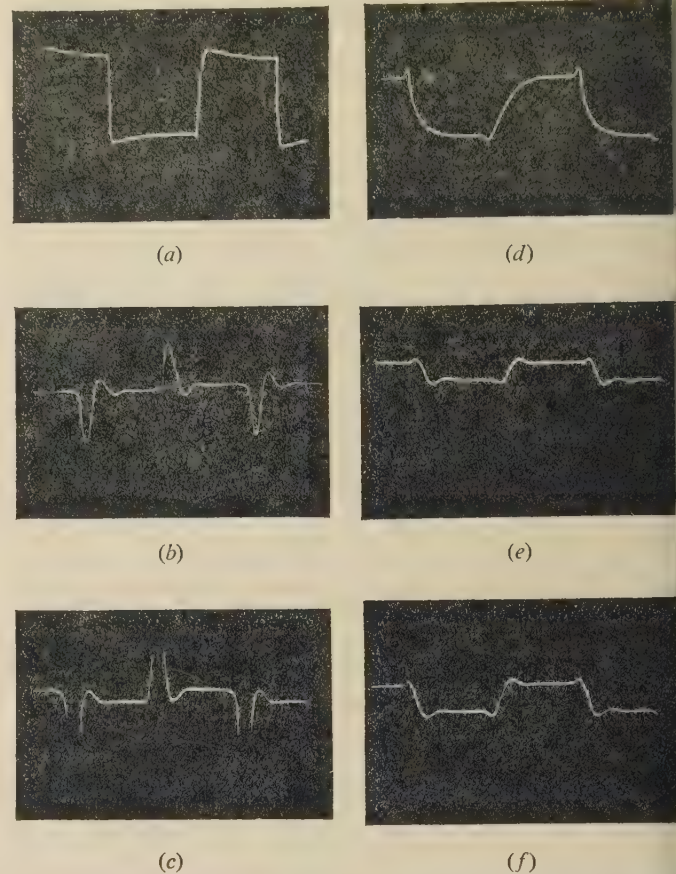
Fig. 5.—Oscillograms taken from an experiment with the amplifier and deflecting coil.

Voltage drive to amplifier is square wave.
(a)–(c) Driving current $i_a(t)$ for increasing amounts of feedback, $M = 0$, $-17 \cdot$ and $-24 \cdot 8$ dB, respectively.
(d)–(f) Flux-producing current $i(t)$ for similar amounts of feedback.

10 000 per second; and, for the purpose of computing these curves, a maximum displacement error of one part in 500 ($d = 500$) is assumed tolerable.

(3.1.4) Experimental Confirmation of the Improvement due to Negative Feedback.

An experiment was made with the amplifier and deflecting coil to confirm that negative feedback could noticeably improve the speed of response. The test waveform was a square wave. A measure of the driving current $i_a(t)$ was obtained by monitoring the voltage across a small resistance in series with the deflecting coil, while a measure of the flux-producing current $i(t)$ was obtained by monitoring the voltage output from the Miller valve (Fig. 2). From Fig. 5, without feedback the waveform of $i_a(t)$ was reasonably square, but because of the rapid changes of slope the waveform of $i(t)$ contained very considerable errors. With increasing amounts of negative feedback, however, large pulses in the waveform of $i_a(t)$ tended to improve the waveform of $i(t)$.

A further experiment was made under more typically variable-velocity conditions. In this case, a square-wave modulating signal was injected into the sweep generator in order to vary the slope of its output waveform, and this generator provided the input voltage drive to the experimental amplifier driving the deflecting coil. Monitors of $i_a(t)$ and $i(t)$ were obtained as previously. The required changes of slope at transitions (4 : 1) were considerably less than those for the square wave, and the errors in $i(t)$ were expected to be correspondingly reduced

(a)   (d)

(b)   (e)

(c)   (f)
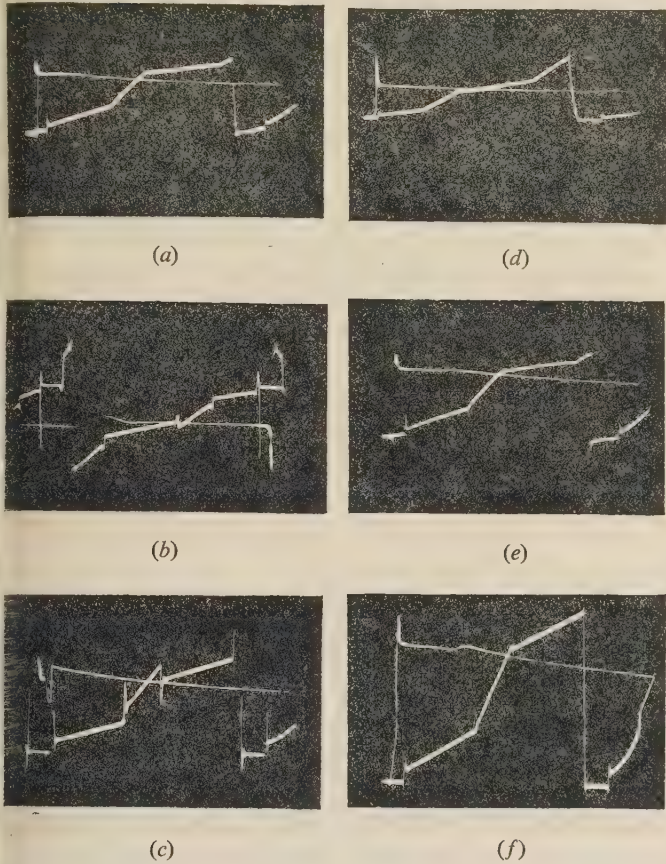
Fig. 6.—Oscillograms taken from an experiment with complete variable-velocity time-base comprising sweep generator, amplifier and deflecting coil.

Voltage drive to sweep generator is square wave.
Corresponding voltage drive to amplifier is modulated sawtooth wave.
(a)–(c) Driving current $i_a(t)$ for increasing amounts of feedback, $M = 0, -17 \cdot 5$ and $-24 \cdot 8$ dB, respectively.
(d)–(f) Flux-producing current $i(t)$ for similar amounts of feedback.

Fig. 6 shows that this was so. In fact, the errors in $i(t)$ were so small (even without feedback, the error was computed to be, at most, one part in 1 500) that improvements due to increasing the amount of negative feedback were hardly noticeable. However, the error-correcting pulses in $i_a(t)$, due to negative feedback, were clearly visible.

(3.2) Transient Analysis for Shunt Damping

For shunt damping (Fig. 8),

$$\frac{\bar{i}}{\bar{i}_a} = \frac{1}{\left(\dfrac{p}{\omega_0}\right)^2 + \dfrac{p}{\omega_0}\left(Q_1 + \dfrac{1}{Q_2}\right) + 1 + \dfrac{r}{R}}$$

where $Q_1 = \omega_0 L/R$, $Q_2 = 1/\omega_0 C_c r$ and $\omega_0^2 = 1/LC_c$

In a form analogous to eqn. (3), with negative feedback, the flux-producing current is given by

$$\bar{i} = \frac{\overline{V_{in}}A_3}{1 - M + \dfrac{r}{R}} \cdot \frac{1}{\left(\dfrac{p}{\Omega_0}\right)^2 + \left(Q_1 + \dfrac{1}{Q_2}\right)\dfrac{1}{\sqrt{\left[(1-M) + \dfrac{r}{R}\right]}}\dfrac{p}{\Omega_0} + 1}$$

$$. \quad . \quad . \quad . \quad (13)$$

where $$\Omega_0 = \omega_0\sqrt{\left[(1-M) + \frac{r}{R}\right]}$$
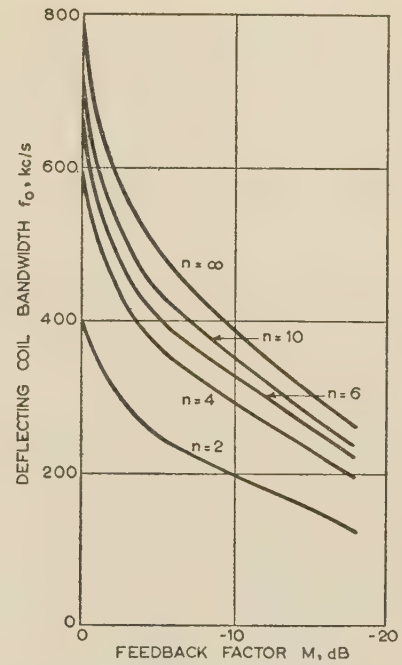
Fig. 7.—Deflecting-coil bandwidth, f_0, necessary for a maximum displacement error of 1/500 of the total scanning length.
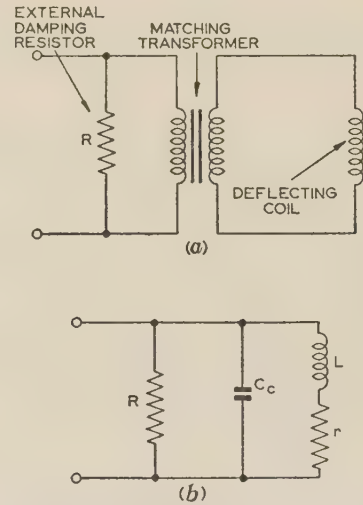
Fig. 8.—Shunt-damped deflecting coil and output-matching transformer.
(a) Exact circuit diagram.
(b) Simplified equivalent diagram.

After comparing eqns. (3) and (13), it will be clear that a simple transform pair of equations is sufficient to convert results for series damping into results applicable to shunt damping. The transform pair is given below:

Series damping	Shunt damping				
M	$M - \dfrac{r}{R}$				
$1/Q_s$	$1/Q_2 + Q_1$	.	.	.	. (14)

During the transient condition, the amplifier delivers a pulse of current which is $M - r/R$ times the transient pulse in the

deflecting coil. Because M is negative, the pulse is of such a sign as to reduce the error current in the coil. The amplifier pulse is greater for shunt damping than for series damping, $M - r/R$ times instead of only M times the transient coil current, owing to the necessity of delivering additional current to the shunt resistance.

It can be easily shown that the shunt-damping resistance R is given by

$$R = \frac{Q_2 r}{1/Q_2 + \sqrt{(1-M)}} \quad . \quad . \quad . \quad . \quad (15)$$

If the maximum tolerable error is given, a formula for the bandwidth f_{-M} of the deflecting coil can be obtained from eqns. (12) and (14):

$$f_{-M} = \frac{d\left(1 - \frac{1}{n}\right)}{2\pi T_s \sqrt{\left[(1-M) + \dfrac{r}{R}\right]}} \quad . \quad . \quad . \quad (16)$$

It has been demonstrated that the transient behaviour of a series-damped coil is similar to that of a shunt-damped coil. Deflecting-coil construction together with considerations of the output-valve matching will, however, clearly indicate which of these two systems is preferable.

(4) OUTPUT-VALVE LOADING

The loading conditions for the two cases—series and shunt damping—are analysed in this Section. For both cases, the peak voltage and peak current required from the output valve are illustrated by means of numerical examples, the parameters of which have been taken from the actual experimental apparatus.

(4.1) Series Damping

The point of difference between series damping and shunt damping lies in the fact that for the latter external damping, in addition to that provided by the deflecting coil, is permissible: for the series case, damping must be provided solely by means of the self-resistance of the deflecting-coil. Now, for a deflecting coil wound in the normal manner with enamelled-copper wire, the damping achieved by the self-resistance and by the hysteresis losses of the highly permeable yoke is insufficient. In order to increase the series resistance, and hence the damping, one method would be to wind the deflecting coil with high-resistance wire (e.g. Eureka). However, this method may lead to difficulties in coil design, but, more important still, the damping achieved thus cannot be varied easily.

A current of approximately sawtooth form is delivered to the deflecting coil by the amplifier A_3 (Fig. 2). If the scan rate is a maximum at the end of the stroke, the voltage required from A_3 is the greatest possible. This voltage, V_s, is given by

$$V_s = LU_1 + I_{pp}r \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where U_1 is the maximum scanning rate, I_{pp} is the peak-to-peak scanning current, and L and r are the reflected parameters of the deflecting coil (Fig. 3).

Eqn. (17) demonstrates that the magnitude of V_s is intimately affected by the magnitude of r. This resistance is given by eqn. (4).

During a transition, the amplifier A_3 delivers M times the error current in the deflecting coil [eqns. (8) and (10)]. Thus, the peak-to-peak current is increased by negative feedback to a value I_{ppfb} given by

$$I_{ppfb} = I_{pp} - M\frac{I_{pp}}{d}$$

where, $1/d$ is the fractional allowable error.

The circuit design is such that d is of the order of 500. Thus, for $-50 < M < 0$, the value of the peak-to-peak current is not appreciably increased by negative feedback.

The following example serves best to illustrate the output requirements:

Example.

Deflecting coils of 2 mH inductance are driven through a 4/1 step-down transformer from a pentode output valve. The resonant frequency of the transformer with valve and deflecting coils connected is 173 kc/s; the coil requires 0·6 amp peak-to-peak current for deflection; the line period $T_s = 10^{-4}$ sec.

From eqn. (4), the series resistance r is given by

$$r = 2 \times 4^2 \times 2 \times 10^{-3} \times 2 \times .173 \times 10^3 (1 - M)^{1/2}$$
$$= 69\,600(1 - M)^{1/2}$$

and, from eqn. (17),

$$V_s = 4^2 \times 2 \times 10^{-3} \times \frac{0·6}{4 \times 10^{-4}} + \frac{0·6}{4} \times r$$
$$= \qquad 48 \qquad\qquad + 10\,400 \text{ volts} \quad (18)$$

Eqn. (18) illustrates the inefficiency of a system employing series damping in order to arrive at the critically damped case of the solution of eqn. (3). In the absence of negative feedback, the output valve is required to deliver up to 10 448 volts at 0·125 amp; only 48 volts appear across the inductive load. If negative feedback is applied, the voltage requirement is even greater.

Series damping has thus been shown to involve difficulties in the construction of the deflection coil; also, the design achieved is not flexible to compensate for changes in values of the circuit elements; further, series damping requires very large driving power.

(4.2) Shunt Damping

Shunt damping is provided by an external shunt resistance R (Fig. 8): this resistance can be varied to compensate for circuit changes. Deflecting coils of high Q-factor can be employed.

Suppose a pentode output valve is employed to drive the deflecting coil through a matching transformer. The output of this valve may be represented by a constant-current generator: thus, the coil damping must be provided entirely by the circuit. On the other hand, the output of a triode may be represented by a constant-current generator shunted by a resistance R_t. If such a valve is used to drive the deflecting coil through a matching transformer, some of the required damping is provided by the valve resistance R_t. For this arrangement, the resistance R_c across the deflecting coil is given by

$$R_c = \frac{RR_t}{R_t - R} \quad . \quad . \quad . \quad . \quad (19)$$

where R is the net damping resistance given in eqn. (13).

The maximum voltage, V_p, required from the output valve is given by

$$V_p = LU_1 + I_{pp}r \quad . \quad . \quad . \quad . \quad (20)$$

Although eqn. (20) for V_p closely resembles eqn. (17) for V_s, the values of V_s and V_p can differ considerably. This is because the resistance r (series case) provides all of the circuit damping and is a function of M: on the other hand, r (shunt case) can be made as small as possible, limited only by the techniques of deflecting-coil construction. Thus, V_p can be very much less than V_s.

For shunt damping, the total peak-to-peak current from the

output valve, $I_{tot(-M)}$, increases with negative feedback due to two components: the first component, I_R, is due to the external damping resistance R (this varies with M); the second component, I_c, is due to the correcting pulse ($M - r/R$ times the error pulse in the coil: Section 3.2):

$$I_R = \frac{V_p}{R} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

and

$$I_c = -\frac{1}{d}I_{pp}\left(M - \frac{r}{R}\right) \quad . \quad . \quad . \quad . \quad (22)$$

$$I_{tot(-M)} = I_{pp} + I_R + I_c \quad . \quad . \quad . \quad . \quad (23)$$

In order to illustrate the effect of applying negative feedback, the following two ratios have been defined:

$$\alpha_I = \frac{I_{tot(-M)}}{I_{pp}} \quad . \quad . \quad . \quad . \quad . \quad (24)$$

where α_I is the fractional increase of output valve current when the amount of feedback is increased from zero to M, and

$$\alpha_\omega = \frac{f_0}{f_{-M}} \quad . \quad . \quad . \quad . \quad . \quad (25)$$

where f_{-M} is the deflecting-coil bandwidth required to produce a given performance with feedback $= -M$; f_0 is the coil bandwidth required to produce similar performance without feedback. Both f_0 and f_{-M} can be obtained from eqn. (16). Here, α_ω is the fractional increase of effective bandwidth when negative feedback is applied. The following example serves to illustrate the output requirements for shunt damping.

Example.

A deflecting coil of 2 mH inductance is driven through a 4/1 step-down transformer from a pentode output valve. The resonant frequency of the deflecting coil with the transformer and valve connected is 173 kc/s: the corresponding Q-factor is 13·9 (i.e. $Q_2 = 13·9$). The deflecting coil requires a peak-to-peak deflecting current of 0·6 amp; the line period $T_s = 10^{-4}$ sec; the maximum allowable error is assumed to be 1/500 of the total scan ($d = 500$).
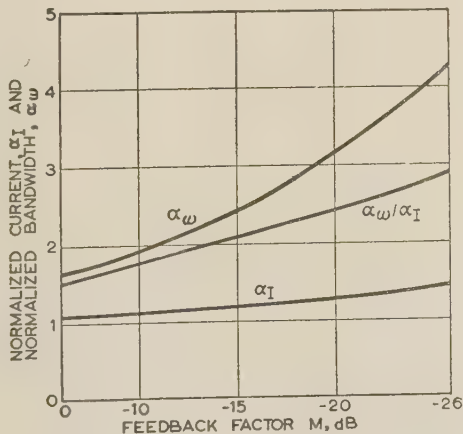
Fig. 9.—Normalized current, α_I, and normalized deflecting-coil bandwidth, α_ω, as functions of feedback factor M.

The variation of the performance with M is illustrated in Fig. 9. As M is made more negative, it can be seen that both the normalized current from the output valve, α_I, and the normalized bandwidth, α_ω, increase monotonically, but the factor by which α_ω is increased greatly exceeds the corresponding factor for α_I, as illustrated by the graph of α_ω/α_I.
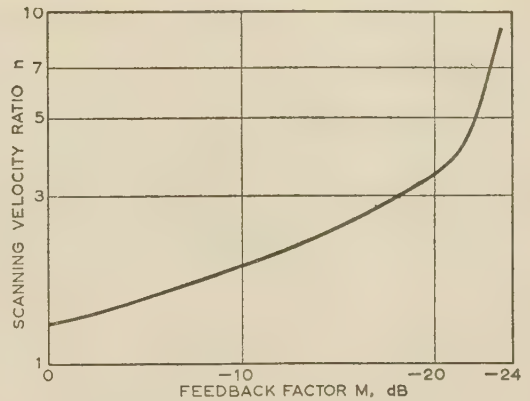
Fig. 10.—Maximum velocity ratio, n, as a function of feedback factor, M, for an assumed maximum tolerable displacement error of 1/500 of the total scan length.

Values of maximum permitted velocity ratio, n, plotted against M, obtained by rearranging eqn. (16), are illustrated in Fig. 10 for an allowed error of 0·2%. Without feedback, the maximum n is 1·3. With 23·5 dB of negative feedback, n may rise to 8·75. The velocity ratio has thus increased by a factor of 6·75, whereas the corresponding increase in output current (Fig. 9) is only 1·38.

(5) CONCLUSIONS

The small overshoot error which is inherent in a magnetic scanner with negative feedback make it eminently suitable for a number of variable-velocity applications which have hitherto been regarded as strictly within the domain of electrostatic scanners. For the magnetic scanner in which damping is achieved by an external resistance placed in shunt with the deflecting coil, the maximum voltage during the stroke is quite small, and an extremely modest power is required from the output valve. Furthermore, this particular system is flexible, allowing for ready adjustment to optimum working conditions.

The present treatment of the magnetic scanner relies very considerably on the results of computation: certain steps of this have, however, been verified by measurements of current waveforms: these show, in fact, that the influence of non-linearities and other factors which have been explicitly ignored in the derivation of this analysis, is negligible in practice.

(6) ACKNOWLEDGMENTS

(7) REFERENCES

(1) DAVIS, R. H.: 'Television Scanning—A Survey', *Electronics*, November, 1935, **8**, p. 30.
(2) BEDFORD, L. H., and PUCKLE, O. S.: 'A Velocity-Modulation Television System', *Journal I.E.E.*, 1934, **75**, p. 63.
(3) CHERRY, E. C., and GOURIET, G. G.: 'Some Possibilities for the Compression of Television Signals by Recoding', *Proceedings I.E.E.*, Paper No. 1401 R, January, 1953 (**100**, Part III, p. 9).
(4) BELL, D. A.: 'Economy of Bandwidth in Television', *Journal of the British Institution of Radio Engineers*, 1953, **13**, p. 447.

(5) RAO, V. N., and SANKAVE SUBRAMANYAN, V.: 'An Expanded Time-base using a Miller Integrator', *Electronic Engineering*, 1955, **27**, p. 273.

(6) SCHROTER, F.: 'Speicherempfang und Differenzbild Fernsehen', *Archiv der Elektrischen Übertragung*, 1953, **7**, p. 63.

(7) GOURIET, G. G.: 'A Method of Measuring Television Picture Detail', *Electronic Engineering*, 1952, **24**, p. 308.

(8) CHANCE, B., HUGHES, V., MacNICOL, E. F., SAYRE, D., WILLIAMS, F. C.: "Waveforms', Radiation Laboratory Series, **19** (McGraw-Hill Book Co., 1949), pp. 281 and 282.

(9) JAEGER, J. C.: 'Introduction to the Laplace Transformation', Methuen Monograph (Methuen, London, 1949).

(10) Reference 8, p. 279.

(11) WESTCOTT, J. H.: 'The Introduction of Constraints into Feedback System Designs', *Transactions of the Institute of Radio Engineers*, Professional Group on Circuit Theory, No. 3, September, 1954.

(12) BODE, H. W.: 'Network Analysis and Feedback Amplifier Design' (D. Van Nostrand, New York, 1945), p. 32.

THE ELECTRICAL BREAKDOWN OF GASEOUS DIELECTRICS AT HIGH FREQUENCIES

By R. R. T. FATEHCHAND, B.Sc.(Eng.), Ph.D., Graduate.

SUMMARY

The continuous-wave breakdown of air and an air/carbon-tetra-chloride mixture in uniform electric fields is examined, at frequencies between 3·5 and 30 Mc/s.

The observations in air indicate that, if the pressure is reduced below atmospheric, the high-frequency breakdown voltage does not fall below the 50 c/s value until the amplitude of positive-ion oscillation has become much smaller than the gap width. This result is explained by a breakdown mechanism which depends both on positive-ion accumulation and on the variation of electron-ionization coefficient with electric field strength. Measurements by an air-stream method of a quantity related to pre-breakdown ionization in the gap appear to be consistent with the suggested breakdown mechanism. The h.f. measurements also exhibit the electron-diffusion type of breakdown, where a simple theoretical relationship must be satisfied for discharge initiation.

A description is given of the effect of a small admixture of carbon tetrachloride vapour on the h.f. breakdown of air at reduced pressure. The vapour concentration was insufficient to affect 50 c/s breakdown. Where the electron-diffusion type of breakdown is normally operative, the h.f. breakdown strength increases by up to 50% when the vapour is present. The increase appears to be caused by electron attachment to vapour molecules.

(1) INTRODUCTION

The paper examines aspects of the high-frequency breakdown of air in uniform electric fields. As distinct from direct current, the high-frequency breakdown initiation processes may be influenced by electron and/or ion oscillation in the discharge gap. In any particular gas, at a fixed field strength, pressure and frequency, the amplitude A_e of electron oscillation is much larger than the amplitude A_+ for positive ions. Thus, if the gap width δ is increased, first A_+ and then A_e will become smaller than δ.

It is known that the h.f. breakdown voltage falls below its 50 c/s value when δ exceeds a critical value which is of the order of A_+. This has been observed[1-3] in air, nitrogen and oxygen at atmospheric pressure and above. The breakdown initiation processes which govern this transition are practically unknown, but it is generally assumed that the effect occurs when A_+ becomes equal to δ, and that it is due to a positive space charge formed in the gap. However, the breakdown voltage observations to be given in Section 3 show that under certain circumstances A_+ must become much smaller than δ, and provide a fuller explanation of the type of breakdown mechanism which is operative. A direct verification that a positive space charge is present in the gap prior to high-frequency breakdown obviously presents great difficulty. Observations made in the present work, however, do go some way in this direction (see Section 4). An ionized air-stream was directed transversely across the discharge gap, and the ions were collected from the emerging air, so that an indication could be obtained of pre-breakdown ionization in the gap.

When the gap width becomes large enough for electrons to oscillate, the breakdown strength abruptly decreases. This decrease has been noticed in many gases.[4-6] Breakdown now occurs when the electron ionization and loss rates balance, and is termed 'diffusion-controlled', since electron diffusion is generally the main source of electron loss. A satisfactory theoretical treatment exists,[7,8] and leads to a simple relationship which must be satisfied at breakdown (see Section 9). The electron loss rate and hence the breakdown strength can be increased by a small d.c. field applied between the electrodes.[9] Electron attachment should have a similar result, and appears to be significant in the results given in Section 5, which discusses the great effect produced by a trace of CCl_4 (carbon tetrachloride) vapour on the h.f. breakdown strength of air when diffusion-controlled breakdown is normally operative. The effect of CCl_4 vapour on the ion current before the h.f. breakdown of air is of interest in this connection, and typical observations will be given in Section 4.

(2) EXPERIMENTAL ARRANGEMENTS FOR BREAKDOWN MEASUREMENTS

The h.f. voltage was measured by a 'dielectric voltmeter'. This instrument[6] relies on the fact that if a slab of dielectric is partly outside a charged parallel-plate condenser, a mechanical force pulls the dielectric in between the plates. The high-frequency r.m.s. voltage can then be determined by measurement of the 50 c/s r.m.s. voltage which produces an equal force on the dielectric. An improved version of the original instrument enabled h.f. voltages to be accurately determined in terms of a 50 c/s voltmeter reading, with a probable error of less than ±1% for voltages above 500 volts r.m.s.

The breakdown measurements were made with sustained oscillations in the frequency range 2·5–30 Mc/s, of negligible harmonic content and with hum modulation less than 1%, so that the peak value of the gap voltage could be accurately determined from the r.m.s. measurement. Fig. 1 shows the discharge chamber and its connection to the dielectric voltmeter. The distance between the dielectric voltmeter and the breakdown gap was only a small fraction of a wavelength in the frequency band used. The electrodes were formed from brass shanks tipped with 1·3 cm diameter platinum discs, which were slightly domed.

During measurements, air was continuously extracted from the chamber and also allowed to enter through a needle valve. The incoming air was passed through a desiccator and a glass-wool filter, so that a stream of clean dry gas circulated in the discharge space. The chamber pressure was determined by mercury gauges. At atmospheric pressure at least, mercury contamination of the discharge chamber did not influence breakdown results. An extensive series of 50 c/s and h.f. measurements were taken with a new chamber, both before and after the connection of a mercury manometer. The two sets of results were essentially the same.

Before breakdown measurements, the electrodes were polished with fine-grade emery paper, washed successively in ethyl alcohol and distilled water, and finally dried. After a few hundred discharges, the breakdown voltages had increased by about 20 volts, and slight electrode burns could be observed. The cleaning procedure was then repeated. For consistent breakdown voltage measurements,[10] a radioactive agent of about 2 millicuries was placed just outside the discharge chamber. Consecutive measurements then had a maximum spread of about 20 volts.

(3) H.F. BREAKDOWN OF AIR

(3.1) Measurements

Fig. 2 shows the results obtained for air at atmospheric pressure, at 50 c/s and at 5–28·5 Mc/s. The peak 50 c/s breakdown voltages agreed within 2% with d.c. measurements, and within 2% with Schumann's[11] results for d.c. breakdown in uniform fields. In this case, 0·6 cm diameter tungsten electrodes were used. All subsequent breakdown measurements were made with the platinum electrodes described in Section 2. For pressures below atmospheric, the peak breakdown voltage $\hat{V}_s$ was plotted against the product $p\delta$ of pressure and gap width, δ being varied for fixed values of pressure and frequency. Fig. 3 shows results obtained at 20 Mc/s, only sufficient experimental points being inserted to indicate that the plotted curves are justified. The 50 c/s Paschen[12] characteristic is also given, and in Fig. 3 is extended down to the minimum voltage for 50 c/s breakdown.

(3.2) Discussion

(3.2.1) Positive-Ion Space-Charge Phenomena.

Figs. 2 and 3 show that, at a fixed frequency and pressure, the h.f. breakdown voltage falls below the 50 c/s value when $p\delta$ exceeds a critical value $(p\delta)_+$. At a fixed pressure, $(p\delta)_+$ decreases as the frequency is increased. This transition will be denoted by T_+, and occurs at the critical gap δ_+. For $p\delta > (p\delta)_+$ all the h.f. breakdown characteristics, except those obtained at 20 Mc/s for pressures below 30 mm Hg (Fig. 3), tend to a common curve which lies 15–20% below the 50 c/s breakdown voltages. The breakdown strength then appears to remain fairly constant up to much higher frequencies.[6] The h.f. breakdown voltage thus tends to a function of the product $p\delta$ of pressure and gap width, so that a Paschen type of law is followed which holds over a large frequency range. It will be shown later that in this region the amplitude A_+ of positive-ion oscillation in the breakdown field is much smaller than δ. Hence the loss of positive ions out of the discharge space will be governed mainly by diffusion. It is to be noted that measurements[13] of the d.c. breakdown of air with an abnormally large positive-ion concentration in the gap prior to breakdown indicate that the reduced V_s also obeys a Paschen type of law.

As discussed in Section 1, at δ_+ the value of A_+ is of interest. To determine A_+ the mobility constant k_+ must be known. Measurements[14,15] for the K+ ion in nitrogen and the N_2^+ ion in nitrogen indicate that for E/p up to 200 volts/cm per mm Hg, a reasonable estimate of k_+ is 3 (cm/sec)-mm Hg per volt/cm. Table 1 compares δ_+ and A_+ for the results of Fig. 2.

It will be seen from Table 1 that A_+ is smaller than δ_+, and the discrepancy becomes greater as the frequency f and $\hat{E}_s/p$ increase. The h.f. breakdown measurements of Lassen[1] and Bright[3] in air and nitrogen at atmospheric pressure show the same tendency. An examination of Ganger's[2] critical gaps for nitrogen, however, indicates that, over a pressure range of 1–30 atm and for differing electrode configurations, δ_+ and A_+ are in reasonable agreement. Here f was low (0·12 Mc/s), and

Table 1
POSITIVE-ION CRITICAL GAPS (AT ATMOSPHERIC PRESSURE)

Frequency	$\hat{E}_s/p$ (at critical gap)	δ_+ critical gap	A_+ oscillation amplitude of positive ion
Mc/s	volts/cm per mmHg	cm	cm
5	85	0·03	0·012
10	93	0·02	0·007
15	98	0·018	0·005
20	107	0·014	0·004
28·5	114	0·011	0·003

$\hat{E}_s/p$ therefore had the comparatively small values of 30–40 volts/cm per mm Hg.

If the present results at reduced pressure are considered, it is found that A_+ becomes much smaller than δ_+, and $\hat{E}_s/p$ attains high values. Fig. 4 shows the increasing discrepancy between A_+ and δ_+ as the pressure is reduced, results obtained at 3·5, 10 and 20 Mc/s being plotted. It will be seen from Figs. 3 and 4 that the critical value $(p\delta)_+$ initially decreases as the pressure decreases or the frequency increases, but finally attains a least value $[(p\delta)_+]_m$ which remains fixed for further pressure or frequency changes. This minimum value is about 5·5 mm Hg-cm, and corresponds to $\hat{E}_s/p \simeq 150$ volts/cm per mm Hg. Thus $\hat{E}_s/p$ tends to a high fixed value, so that A_+ becomes constant for a particular frequency, while δ_+ tends to vary inversely as the pressure.

The following explanation of the observations is based on the nature of the relationship between α/p and E/p, shown for air[16] in Fig. 5(a), where α is the electron-ionization coefficient. Consider the number $\hat{n}$ of electrons produced in an electron avalanche starting from the neighbourhood of the temporary cathode when the applied field is near its peak value. In an undistorted field, $\hat{n} = \exp(\hat{\alpha}_0\delta)$, where $\hat{\alpha}_0$ is the value of α at the peak value of the applied breakdown field $(\hat{E}_s)_{hf}$. If the field is distorted by space charge, however, $\hat{\alpha}$ will not be constant across the gap, and $\hat{n}$ will be given by $\exp\left(\int_0^\delta \hat{\alpha}\,dx\right)$. Since $\exp(\alpha\delta)$ varies extremely rapidly with field strength, most of the contributions to the positive-ion space charge will be produced by electron avalanches which cross the gap when the applied field is near a maximum. It also follows that breakdown will occur when the applied field is near its peak value. If the positive space charge distorts the electric field across the gap, so as to increase $\hat{n}$ above its value in the undistorted field, an unstable condition might be expected, finally leading to a discharge. In this case, therefore, the breakdown voltage will be reduced. If space charge cannot change $\hat{n}$, however, the h.f. breakdown voltage should equal the direct-current or 50 c/s value. For a uniform space-charge distribution, the total field will vary linearly across the gap, as indicated in Fig. 5(b). In Fig. 5(c) the variation of $\hat{\alpha}$ across the gap has been plotted for the three main regions of the α/p, E/p characteristic. When α/p varies more rapidly than E/p, which is the case[17] up to $E/p = 130$, the value of $\int_0^\delta \hat{\alpha}\,dx$, and thus $\hat{n}$, is increased by field distortion. When E/p is between about 130 and 270, the relationship between α/p and E/p is linear, and the integral is unchanged by field distortion. At still higher values of E/p, α/p varies more slowly than E/p, and the integral is actually decreased.

In the observations given here, the minimum value of $(p\delta)_+$ is about 5·5 mm Hg-cm, and the corresponding value of $\hat{E}_s/p$ is 150 volts/cm per mm Hg. This agrees well with the value of
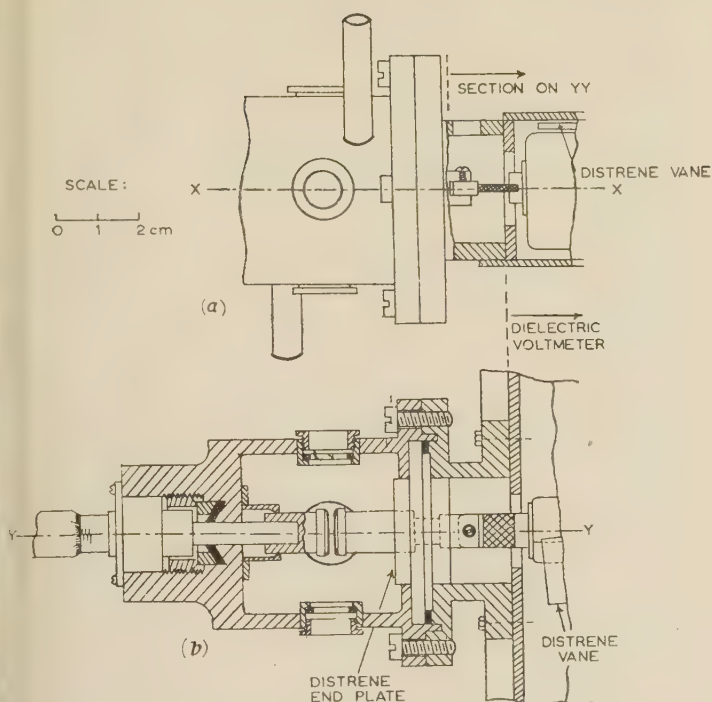
Fig. 1.—Discharge chamber.

(a) Elevation.   (b) Plan view: section on XX.
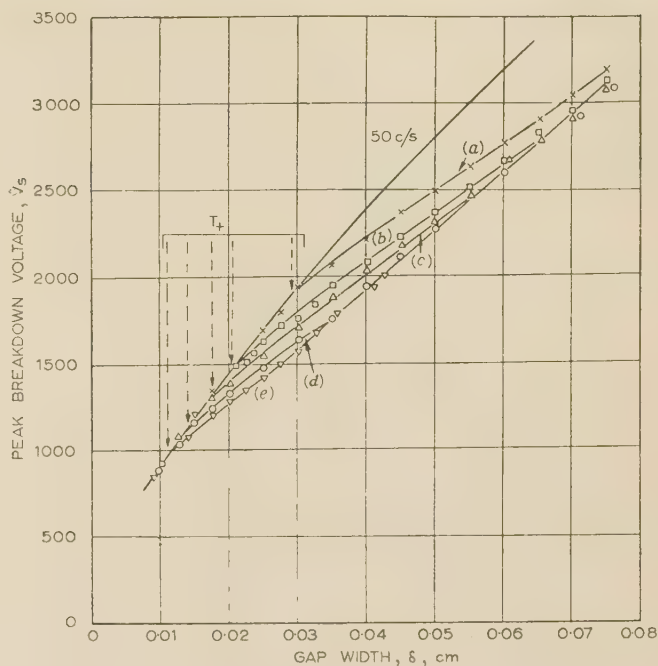
Fig. 2.—H.F. breakdown of air at atmospheric pressure.

(a) $f = 5$ Mc/s.   (c) $f = 15$ Mc/s.
(b) $f = 10$ Mc/s.   (d) $f = 20$ Mc/s.
(e) $f = 28 \cdot 5$ Mc/s.
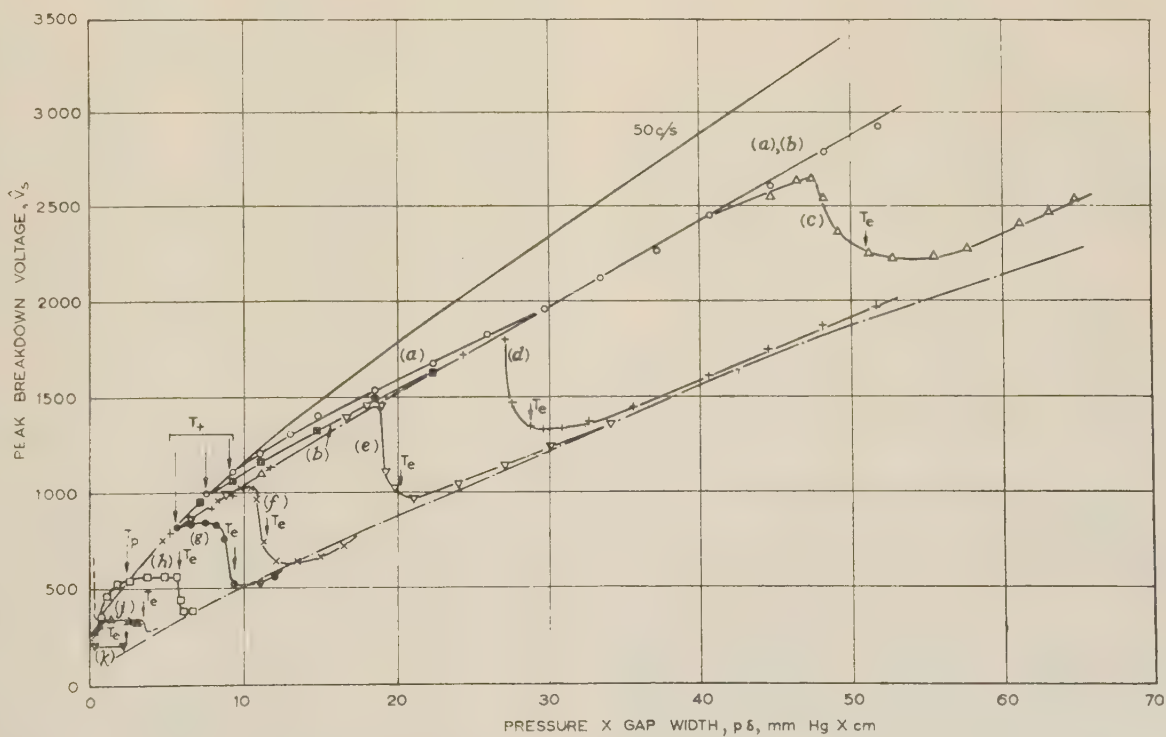
Fig. 3.—20 Mc/s breakdown of air at fixed pressures.

(a) 738 mm Hg.   (f) 30 mm Hg.
(b) 369 mm Hg.   (g) 22 mm Hg.
(c) 185 mm Hg.   (h) 12 mm Hg.
(d) 93 mm Hg.   (j) 6 mm Hg.
(e) 60 mm Hg.   (k) 3·5 mm Hg.
— · — Theoretical electron diffusion breakdown curve.
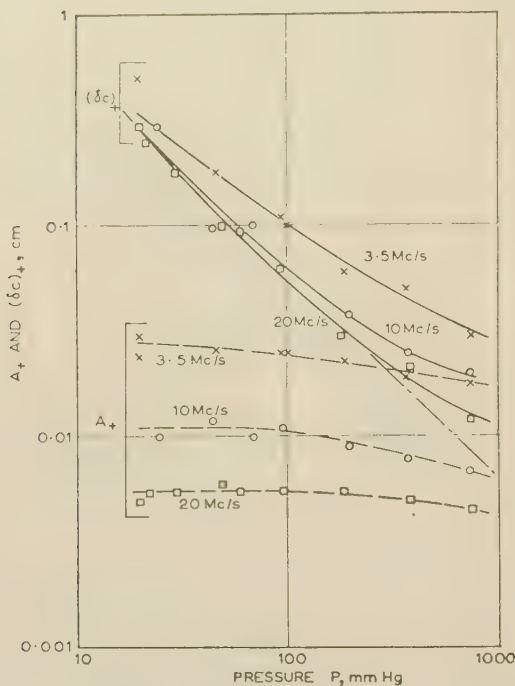
Fig. 4.—Variation of critical gap width δ_+ and amplitude of positive-ion oscillation A_+ with pressure, for air.

—·— $p\delta = 5\cdot5$.
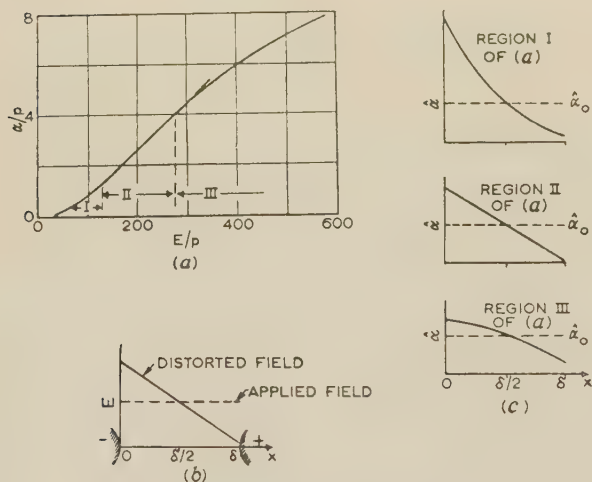
Fig. 5.—Field distortion by space charge.

(a) Variation of α/p with E/p for air (Townsend).
(b) Field distortion with uniformly distributed positive space charge.
(c) Effect of space charge on electron ionization coefficient $\hat{\alpha}$, applied field at peak value.

E/p above which field distortion cannot be expected to increase $\hat{n}$. If $p\delta$ is reduced below $5\cdot5$ mm Hg-cm, E/p increases above 150 volts/cm per mm Hg, so that the h.f. breakdown voltage will not fall below the 50 c/s value, even although the amplitude A_+ of positive-ion oscillation in the breakdown field is smaller than the gap width δ, and a positive-ion space charge can accumulate. It follows that at low pressures or high frequencies, where $(p\delta)_+$ tends to be low but still greater than $5\cdot5$ mm Hg-cm, a large positive-ion accumulation must exist before gap ionization can be appreciably increased by field distortion. Thus A_+ must become much smaller than δ_+ before the h.f. breakdown voltage

will fall below the 50 c/s value. At high pressures or low frequencies, on the other hand, $(p\delta)_+$ is high and $\hat{E}_s/p$ is low, so that breakdown should set in for small initial positive-ion accumulations. Hence, A_+ may be expected to be in agreement with δ_+. This seems to be confirmed by the present results, and by the previously discussed measurements of other experimenters in air and nitrogen at atmospheric pressure and above.

The above breakdown mechanism was suggested by Varney's[17] analysis of the influence of a positive-ion space charge on d.c. discharge initiation when an abnormally large number of primary electrons are present in the gap. Although in this case the d.c. breakdown voltage is generally reduced, this is not always so, and it can be increased[18] if the experiments are made near the minimum breakdown voltage, where E/p is very high and α/p varies more slowly than E/p, i.e. in region III of Fig. 5(a). It will be noticed that at a pressure of 12 mm Hg the 20 Mc/s breakdown characteristic of Fig. 3 exhibits a small rise above the 50 c/s curve. The rise is found at gaps too small for electron oscillation effects to be present, and at values of $p\delta$ near the 50 c/s breakdown minimum. This occurrence is presumably associated with the space-charge condition of Fig. 5(c), where $\hat{n}$ is decreased. A similar explanation may possibly cover the increase in breakdown strength of hydrogen at high frequencies, noticed by Fuchs.[19]

(3.2.2) Electron Oscillation Effects.

The abrupt nature of the transition T_e, above which electrons can oscillate in the gap, is exhibited in Fig. 3. At T_e the amplitude A_e of electron oscillation in the breakdown field is equal to δ, taking[20] a value for the electron mobility constant k_e of 527 (cm/sec)-mm Hg per volt/cm. For gap widths above the T_e transition, the h.f. breakdown voltage $(\hat{V}_s)_{hf}$ is a function of $p\delta$, so that a Paschen type of law is again obeyed. The equation given in Section 9 indicates that this should be the case for the diffusion-controlled mode of breakdown. The values of $(\hat{V}_s)_{hf}$ obtained from this equation are plotted in Fig. 3, and are in fair agreement with the observations in the diffusion-controlled breakdown region. Measurements of R by Townsend and Tizard,[21] and measurements of α/p by Sanders,[22] were used.

The electron effect first noticed by Pim[6] occurs in the results of Fig. 3. Here electrons which return across the gap as the field reverses can make a significant contribution to discharge space ionization, and thus reduce the rate of increase of h.f. breakdown voltage with gap width. The effect is generally a minor one in the present experiments, and in Fig. 3 is only indicated (T_p) for the gas pressure of 12 mm Hg.

(4) ION FORMATION PRIOR TO H.F. BREAKDOWN

(4.1) Experimental Arrangements

The following, an extension of earlier work[23] by the author, is an attempt to verify that positive ions oscillate in the gap prior to the breakdown of air at high frequency. The arrangement is shown schematically in Fig. 6. A nozzle directed a stream of ionized air transversely across the discharge gap, which consisted of two slightly domed tungsten electrodes of $0\cdot6$ cm
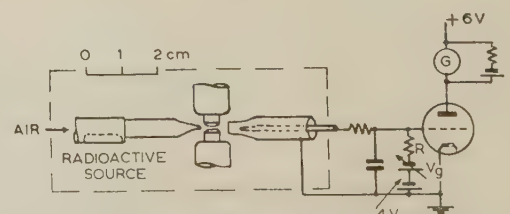
Fig. 6.—Air-stream method for measurement of ion current.

diameter. The ions which emerged were collected by a negatively-charged wire, an electrometer valve being used to measure the voltage which the ion current developed across a high resistance R. Air turbulence was minimized by tapering the nozzle orifice and ion-collector assembly in the neighbourhood of the discharge gap. A stream of dry clean air was pumped through the nozzle, and before emerging passed over a radioactive agent of about 2 millicuries. Any h.f. voltage induced on the wire probe was eliminated from the electrometer valve grid by an RC filter. The valve grid and probe were biased to -4 volts. The variable voltage V_g was adjusted to compensate for the voltage drop V_i produced in R by the ion current I, using the galvanometer in the valve anode circuit as a null-indication device. The probe voltage was thus maintained constant, and I was given by V_i/R.

The initial breakdown invariably occurred across the electrode centres. The static air pressure in the gap remained at atmospheric, since the application of the air stream had no effect on the breakdown voltage values. With an air flow of 10–15 litres/min through the nozzle, and for gap widths greater than 0·025 cm, steady ion-current readings could be obtained up to very near breakdown.

(4.2) Measurements

Fig. 7 exhibits the variation of ion current with applied voltage, for fixed gap widths. All the curves of (c) and (d), and the
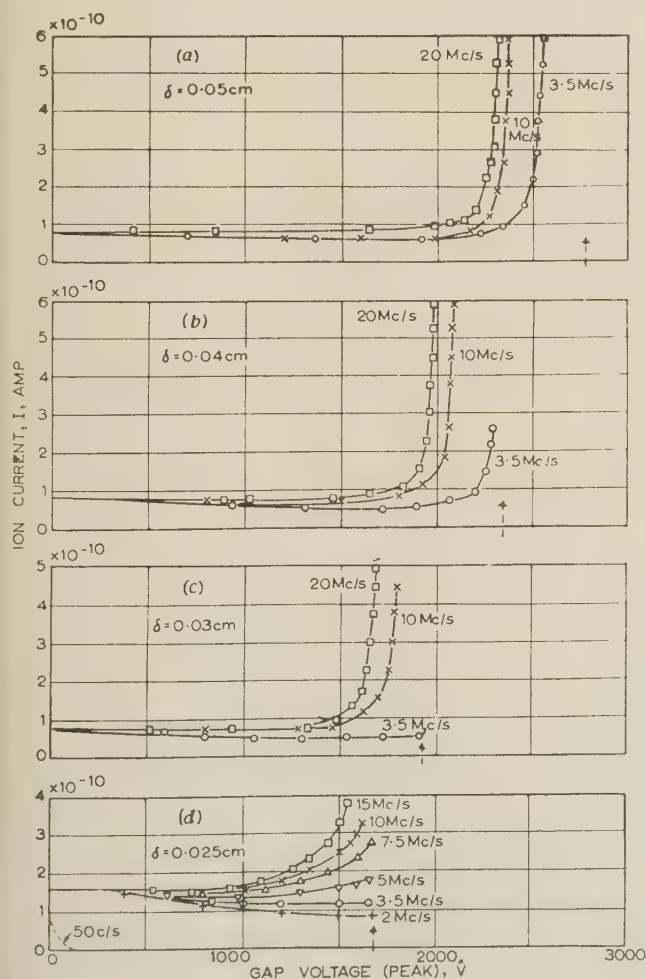
Fig. 7.—Pre-breakdown ion currents at fixed gap widths, for air.
Vertical arrows indicate 50c/s peak breakdown voltages.
(a), (b), (c) at air flow of 10 litres/min.
(d) at air flow of 15 litres/min.

3·5 Mc/s curve of (b) terminate in the last steady ion-current value observed before the onset of breakdown. The other curves increased before breakdown to currents of $10-15 \times 10^{-10}$ amp, but the high-current regions have been omitted. However, in these regions the current rise was extremely rapid, and all curves terminate at voltages very close to discharge initiation. With no applied gap voltage, the current was produced solely by the irradiation of the air stream.
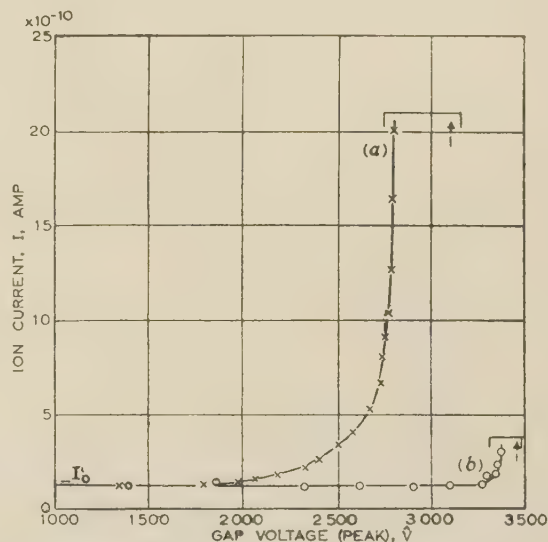
Fig. 8.—Influence of CCl_4 vapour on pre-breakdown ion current for air.
Vertical arrows indicate 50c/s peak breakdown voltages.
I_0' is ion current for $V = 0$. Air flow = 15 litres/min.
Gap width = 0·06 cm. Frequency = 3·5 Mc/s.
(a) Air. (b) Air/CCl_4, with 1% of mixture CCl_4 vapour.

Fig. 8 shows a typical modification of ion-current characteristic produced by the introduction of a small quantity of CCl_4 vapour into the air stream before it left the nozzle. The CCl_4 molecules formed between 1 and 2% of the air/vapour mixture, sufficient to increase the 50c/s breakdown voltage[24] by 13%.

(4.3) Discussion

At 50c/s, the ion current I falls rapidly to zero as the gap voltage $\hat{V}$ is increased. Use of the appropriate mobility-constant values (Section 3) indicates that at 50c/s, and as $\hat{V}$ is increased, the amplitudes of oscillation, A_+ and A_e, of ions and electrons will soon become much greater than the gap width δ, so that very few charged particles can be expected to enter or leave the gap via the air stream. At the gap widths considered, and for the high frequencies used in the measurements, A_e will still become greater than δ at quite small values of $\hat{V}$, but this will not be the case for negative or positive ions. Since at high frequencies there is only a small initial reduction in I as $\hat{V}$ increases, I must consist mainly of ions.

The ion-current curves should be considered in conjunction with Fig. 2, which shows that at a fixed δ the h.f. breakdown voltage falls below the 50c/s value when the frequency exceeds a critical value f_c which diminishes as δ increases. Curves (c) and (d) of Fig. 7 indicate that the neighbourhood of f_c is associated with a rapid rise of I just before breakdown, and that as f increases above f_c this rise becomes increasingly pronounced. As shown in (c) and (d) of Fig. 7 there is a pre-breakdown increase of I for frequencies less than f_c. This increase is unlikely to be due to electrons, which should be lost to the

electrodes, and production of positive ions will therefore be considered. Table 1 shows that, at the critical frequencies for (c) and (d) of Fig. 7, the amplitude of oscillation A_+ of the N_2^+ ion in the breakdown field will be considerably less than δ. Thus, for frequencies somewhat below f_c, positive ions produced in the discharge space should be able to oscillate and hence escape in the air stream. This is clearly shown in (d), where $f_c = 7 \cdot 5$ Mc/s, and $A_+ = \delta$ at a frequency of $2 \cdot 5$ Mc/s. It thus appears that accumulation of positive ions in the gap is not a sufficient criterion for the h.f. breakdown voltage to decrease below the 50 c/s value. This conclusion was reached in Section 3.2.1 by consideration of the dependence of the h.f. breakdown voltage on various parameters. It can be seen that when f_c is exceeded, the breakdown voltage falls as the pre-breakdown current increases. This presumably indicates that increased positive space charge correlates with diminished breakdown strength. The effect of the space charge will increase with frequency until A_+ has become so small that the rate of loss of ions is governed mainly by diffusion.

Interesting consequences follow from the introduction of CCl_4 vapour into the air stream. Fig. 8 shows that the ion-current curve now exhibits no rise until the gap voltage reaches a value very close to breakdown. Both the h.f. and the 50 c/s breakdown strengths increase when the vapour is introduced, the h.f. breakdown voltage approaching the increased 50 c/s value. The vapour thus tends to inhibit the positive-ion space-charge mode of breakdown. Consideration of Fig. 7 indicates that the effect of the vapour on ion current and breakdown voltage is similar to that of frequency reduction at a fixed gap width. This suggests a fall in effective positive-ion accumulation in the discharge space, such as might occur if electrons attached to form negative ions. The possibility of electron attachment is further discussed in the following Section.

(5) BREAKDOWN OF AN AIR/CARBON-TETRACHLORIDE MIXTURE

(5.1) Measurements

The experiments were made at a pressure of 98 mm Hg, which enabled both the positive-ion and electron oscillation regions to be conveniently studied. A mixture of air and CCl_4 vapour was drawn into the discharge chamber through the needle valve mentioned in Section 2. Fig. 9 shows the breakdown voltage
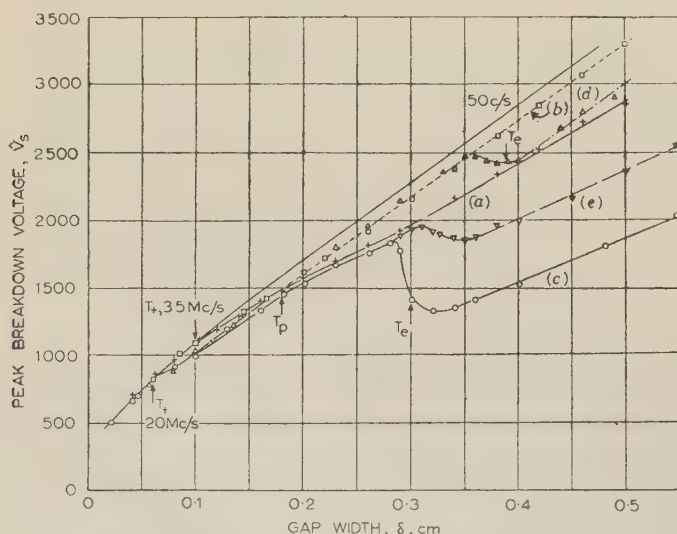
Fig. 9.—Breakdown of air/CCl_4 at 98 mm Hg pressure.

(a) Air; $f = 3 \cdot 5$ Mc/s.   (b) Air/CCl_4; $f = 3 \cdot 5$ Mc/s.
(c) Air; $f = 20$ Mc/s.   (d) Air/CCl_4; $f = 20$ Mc/s.
e) Air; superimposed d.c. field of 300 volts/cm, $f = 20$ Mc/s.

measurement at 50 c/s and at $3 \cdot 5$ and 20 Mc/s, plotted as a function of $p\delta$. The effect of a small superimposed unidirectional field $E_{dc} = 300$ volts/cm is also indicated.

The addition of CCl_4 changed 50 c/s breakdown voltages very slightly, probably by less than 1%. From this it can be estimated[24] that probably fewer than 1 in 500 molecules of the mixture were CCl_4.

(5.2) Discussion

The following points are of interest:

(a) In no case was the positive-ion critical gap δ_+ changed by the addition of the vapour. For gap widths greater than δ_+, the h.f. breakdown characteristics tended to remain parallel to the 50 c/s curve, and at a comparatively small voltage below it.

(b) Carbon tetrachloride produced a large increase in the h.f. breakdown voltage in the region where the electron-diffusion type of breakdown would normally have occurred. The transition T_p completely disappeared, but there was a region similar to the electron oscillation transition T_e observed in air.

The addition of CCl_4 did not appreciably affect 50 c/s breakdown, and the electron mobility constant k_e was apparently unchanged, so that the electron-ionization coefficient α was presumably unaffected. The h.f. increase in the breakdown strength of the air/vapour mixture might, however, have been produced by an increased electron loss rate due to electron attachment. The attachment parameter η has been measured[25] in CCl_2F_2 (Freon). Similar attachment processes occur[26] in CCl_2F_2 and in CCl_4, and in CCl_2F_2 the 50 c/s and h.f. breakdown strengths have been found to be equal.[2,3] The parameter η is defined as the mean attachments of an electron per centimetre in the field direction, and is thus analogous to α. In CCl_2F_2, for small E/p, $\eta > \alpha$, so that there is a high probability of electron attachment. As E/p is increased above about 120 volts/cm per mm Hg, η falls from a value approximately equal to α to very low values when $E/p > 220$. It is reasonable to assume that the foregoing results for CCl_2F_2 will also apply to CCl_4 vapour, and it will be further assumed that a similar relationship between η and E/p will hold for an air/CCl_4 mixture.

In the h.f. breakdown region where electrons oscillate in the gap, electron capture by CCl_4 molecules becomes probable when the electric field is small. Thus, if an electron normally makes a large number of oscillations before diffusing out of the gap, the change of capture will be high. This could explain the greatly increased h.f. breakdown strength produced by the addition of the vapour to air. It is significant that in the electron-diffusion breakdown region the effect of CCl_4 is similar to that of a small d.c. stress superimposed on the h.f. field, as shown in Fig. 9. The d.c. field increases the rate of loss of electrons out of the discharge space.[9] Electron attachment would also explain the absence of the transition T_p in the 20 Mc/s air/CCl_4 results, since electrons which arrive at the temporary anode as the field is reversing will tend to attach. In the positive-ion space-charge region, CCl_4 produces a small increase in the h.f. breakdown strength. At the E/p values which exist in this region, η is still appreciable, so that the positive space charge is presumably reduced by the presence of negative ions and the breakdown voltage consequently increased. The vapour has no effect in the neighbourhood of the T_+ transition. Here the h.f. breakdown strength is independent of space-charge formation.

Finally, it is of interest to consider some results obtained by the author for the h.f. breakdown of argon at a pressure slightly above atmospheric, the gas being contaminated by air. In the electron-diffusion breakdown region, the breakdown strength was extremely sensitive to the proportion of air present in the parent gas, whereas at 50 c/s the influence of the impurity was much less marked. This effect was possibly produced by the attachment of

low-energy electrons to oxygen molecules, and therefore similar to the phenomena discussed previously.

(6) CONCLUSIONS

Evidence is presented that a positive-ion space charge is conducive to the lowering of the h.f. breakdown voltage of air below the 50 c/s value, and that the nature of the relationship between α/p and E/p decides the influence of the space charge on breakdown strength. It is shown that this type of h.f. breakdown is similar to d.c. breakdown of a special kind. The type of d.c. breakdown considered is one in which a large positive space charge is produced in the gap by intense ultra-violet irradiation of the cathode or other means, and has been investigated by earlier workers.

An examination is made of the effect on the high-frequency breakdown strength of air produced by the addition of a small quantity of CCl_4 vapour. The large increase in electric strength, when electrons can normally oscillate in the gap, is reasonably explained by an electron-attachment loss mechanism.

(7) ACKNOWLEDGMENTS

(8) REFERENCES

(1) LASSEN, H.: 'Frequenzabhangigkeit der Funkenspannung in Luft', Archiv für Elektrotechnik, 1931, 25, p. 322.
(2) GANGER, B.: 'Der Hochfrequenzdurchschlag verdichteter Gase', ibid., 1943, 37, p. 267.
(3) BRIGHT, A. W.: 'Corona and Breakdown at Frequencies up to 12 Mc/s', British Electrical and Allied Industries Research Association, 1950, Technical Report Reference L/T229.
(4) GILL, E. W. B., and DONALDSON, R. H.: 'The Sparking Potential of Air for High-Frequency Discharges', Philosophical Magazine, 1931, 12, p. 719.
(5) GILL, E. W. B., and VON ENGEL, A.: 'Starting Potentials of Electrodeless Discharges', Proceedings of the Royal Society, A, 1949, 197, p. 107.
(6) PIM, J. A.: 'The Electrical Breakdown Strength of Air at Ultra-High Frequencies', Proceedings I.E.E., Paper No. 802 R, March, 1949 (96, Part III, p. 117).
(7) HOLSTEIN, T.: 'Initiation of High Frequency Gas Discharges', Physical Review, 1946, 69, p. 50.
(8) BROWN, S. C.: 'Breakdown in Gases, Alternating and High-Frequency Fields', Handbuch der Physik Band XXII. Gasentladungen 11 (Springer-Verlag, Berlin, 1956), p. 531.
(9) VARELIN, L. J., and BROWN, S. C.: 'Microwave Determinations of Average Electron Energies and the First Townsend Coefficient in Hydrogen', Physical Review, 1950, 79, p. 947.
(10) LOEB, L. B.: 'Fundamental Processes of Electrical Discharge in Gases' (John Wiley and Son, New York, 1939).
(11) SCHUMANN, W. O.: 'Elektrische Durchbruch-feldstarke von Gasen' (Springer-Verlag, Berlin, 1923), p. 25.

(12) PASCHEN, F.: 'Über die zum Funkenübergang in Luft, Wasserstoff, und Kohlensaure bei verschiedenen Brücker erforderliche Potentialdifferenz', Annalen der Physik, 1889, 37, p. 69.
(13) WHITE, H. J.: 'The Variation of Sparking Potential with Intense Ultraviolet Illumination', Physical Review, 1935, 48, p. 113.
(14) HERSHEY, A. V.: 'Measurements of the Mobility of Potassium Ions at High Field Intensity and Low Pressure', ibid., 1939, 56, p. 908.
(15) MITCHELL, J. H., and RIDLER, K. E. W.: 'The Speed of Positive Ions in Nitrogen', Proceedings of the Royal Society, A, 1934, 146, p. 911.
(16) TOWNSEND, J.: 'Electrons in Gases' (Hutchinson, 1948), p. 106.
(17) VARNEY, R. N., WHITE, H. J., LOEB, L. B., and POSIN, D. Q.: 'The Role of Space Charge in the Study of the Townsend Ionization Coefficient and the Mechanism of Static Space Charge', Physical Review, 1935, 48, p. 818.
(18) SEITZ, W., and FUCHS, W.: 'Zundspannungserhöhung durch Ultravioletbestrahlung', Naturwissenshaft, 1936, 24, p. 346.
(19) FUCHS, W., GRAF, L., MUES, G., and MÜLLER, H. G.: 'Zundung der elektrischen Gasentladung im homogenen Wechselfeld', Zeitschrift für Physik, 1956, 145, p. 1.
(20) NIELSON, R. A., and BRADBURY, N. E.: 'Electron and Negative Ion Mobilities in Oxygen, Air, Nitrous Oxide and Ammonia', Physical Review, 1937, 51, p. 69.
(21) TOWNSEND, J. S., and TIZARD, H. T.: 'The Motion of Electrons in Gases', Proceedings of the Royal Society, A, 1913, 88, p. 336.
(22) SANDERS, F. H.: 'Measurement of the Townsend Coefficient for Ionization by Collision', Physical Review, 1933, 44, p. 1020.
(23) FATEHCHAND, R.: 'Positive-Ion Formation in Air Prior to High-Frequency Breakdown', Nature, 1951, 167, p. 566.
(24) RODINE, M. T., and HERB, R. G.: 'Effect of CCl_4 Vapour on the Dielectric Strength of Air', Physical Review, 1937, 51, p. 508.
(25) HARRISON, M. A., and GEBALLE, R.: 'Simultaneous Measurement of Ionization and Attachment Coefficients', ibid., 1953, 91, p. 1.
(26) WARREN, J. W., HOPWOOD, W., and CRAGGS, J. D.: 'On Dissociation Processes in Certain Gases of High Dielectric Strength', Proceedings of the Physical Society B, 1950, 63, p. 180.

(9) APPENDIX

Provided that certain restrictions are satisfied, Holstein[7] has shown that the following condition holds for electron-diffusion breakdown between electrodes at a separation δ, in a gas at pressure p:

$$(p\Lambda)^2 = 0\cdot025\frac{R}{\frac{(E_s)_{hf}}{p}\frac{\alpha}{p}}$$

where Λ is the 'characteristic diffusion length', and is equal to δ/π in the present experiments. R is the ratio of electron to molecular temperature, $(E_s)_{hf}$ the r.m.s. value of the h.f. breakdown field, and α the electron-ionization coefficient. Both R and α/p are determined for a value of E/p equal to $(E_s)_{hf}/p$.

621.385.032.213.13

THE CONDUCTIVITY OF OXIDE CATHODES

Part 2.—Influence of Ion Movements on Matrix Resistance

By G. H. METSON, M.C., Ph.D., M.Sc., B.Sc.(Eng.), Associate Member.

SUMMARY

The present Part continues the study of ion movements in an oxide matrix enclosed between two heated core pieces. Ions are divided into two classes—residual gas ions (probably carbon monoxide) resulting from imperfections of processing technique, and oxygen ions resulting from electrolysis of the oxide matrix. The different influences of the two classes of gas on matrix conductivity are distinguished by characteristic reactions, which are examined in both d.c. and a.c. fields. The effect of electrolytic oxygen on matrix resistance may be profound, and the action is probably a fundamental aspect of cathode operation.

(1) INTRODUCTION

(1.1) Purpose of Present Part

In the first Part of the paper evidence was assembled to show that non-uniformity of resistance may arise in an oxide cathode under current load, and this phenomenon was explained in terms of movement of negative ions of residual gas in the vacuum pores of the granular (BaSr)O matrix. It is proposed in Section 2 of the present Part to extend the study of residual gas behaviour in the matrix, and in Section 3 to introduce the destructive oxygen phenomena which arise from matrix electrolysis.

The work described in the present Part was carried out at the conventional operating temperature of $1\,020°\,K$, and the experimental vehicle was the S-type assembly described in detail in Part 1: its specification is set out below for reference.

> Cores: Active nickel or pure platinum.
> Matrix: Co-precipitated equimolar barium-strontium oxide.
> Matrix density: About $1\cdot0$.
> Matrix thickness: 150μ.
> Matrix area: $0\cdot50\,cm^2$.

The assemblies were vacuum processed to the standard schedule detailed in Table 1 of Part 1.

(1.2) Two Classes of Matrix Gas

Two clearly distinctive classes of gas can be recognized within the oxide matrix—one whose existence is independent of the passage of current and the other which is fundamentally dependent on current flow. The first class is described as residual gas and the second as electrolytic oxygen arising from electrolysis of the basic oxide matrix.

Residual gas is an essentially technological phenomenon resulting from imperfections in the vacuum and materials processing. The gas resides in the hollow pores, and its normal state is that of a negative ion: it has therefore much freedom of movement under thermal or electric forces. The exact nature of the gas will not concern us here, but there is a strong probability that it is carbon monoxide.*

* A mass-spectrometric analysis of the gas driven out of the overheated matrix of an S-type valve showed it to be mainly carbon monoxide. The only gas observed in quantity in receiving pentodes made at the Post Office Research Station is likewise carbon monoxide.

The barium-strontium oxide matrix can be made to show an appreciable electrolytic conductivity at $1\,020°\,K$. Under electric stress the S-type valve will therefore experience a continuous discharge of barium and/or strontium metal at the cathodic core and oxygen at the anodic core. Under certain conditions this electrolytic oxygen can be retained within the matrix pores, and its influence on matrix resistance is a major object of study in the paper.

(1.3) Gas Detection in the S-Type Assembly

It was shown in Part 1 of the paper that a standard S-type valve with active nickel cores and the dimensions set out in Section 1.1 may have any resistance between 40 and 100 ohms immediately after manufacture. Operation of the valve at $1\,020°\,K$, however, always results in a steady decrease of resistance to a constant value of 16 ± 2 ohms, which is thereafter maintained indefinitely. This form of behaviour has been observed without exception on some 50 samples, and the low stable state will henceforth be described as the characteristic resistance, R_0. It was further shown in Part 1 that, once the R_0 state has been achieved, the only way in which the resistance can again be increased is by deliberate introduction of a suitable gas. A gas-free matrix is thus identified with the R_0 state and any deviation from it is taken as a measure of gas within the pore system.

The significance of the characteristic state R_0 will be examined in a later Part of the paper, where it will be shown that its inherent stability is due to a space-charge phenomenon. The concept has been introduced here to provide a fixed datum to which may be referred changes of resistance R_d consequent on the presence of free gas ions in the pore system.

(1.4) Chemical Considerations

Before proceeding to the experimental work it will be useful to consider two points of a chemical nature that will arise in the course of the work. The first concerns the possibility of chemical interaction in the S-type valve between the core metal and the gas. Two distinct conditions can be visualized—the core metal may be chemically active and act as a sink to gas, or it may be inert, reflecting all gas back into the matrix. Under current load the active case will steadily reduce the gas within its matrix to extinction; the inert case will maintain its level of residual gas constant and build up its concentration of electrolytic oxygen on a Faradaic basis. The resistance of the active case will therefore fall and that of the inert case rise with operating time.

The chemically active case is realized in practice by employing the nickel alloy commonly used in the manufacture of commercial oxide cathodes. The alloy contains a small proportion of magnesium metal, and at $1\,020°\,K$ is readily able to act as sink to both residual gas and electrolytic oxygen. Such cores were described in Part 1, where it was shown that the resistance of the device moves inexorably to the low R_0 state where the resistance is invariant with time and current. Such an active system is clearly unsuited for work on the effects of electrolysis. The inert

system is achieved by using pure platinum cores, which are unaffected chemically by either oxygen or carbon monoxide. It is this system which was used exclusively in the electrolytic investigations reported in Section 3.

The second point requiring consideration is the manner in which electrolytic conductivity develops in the S-valve matrix. It will be shown that a valve can be brought to its highest possible level of electronic conductivity—synonymous with the R_0 state—without developing any observable level of electrolytic conductivity. The two conductivities have, in fact, no obvious relationship. Experience seems to indicate that the essential factor in the development of electrolytic conductivity is thermal sintering of the porous matrix. Valves prepared according to the standard thermal schedule* detailed in Table 1 of Part 1 have, for example, a negligibly low level of electrolytic conductivity, and any resistance deviations from R_0 in such valves are due wholly to residual-gas action. If, however, the thermal schedule of Table 1 is modified to include an adequate period of sintering at higher temperature, two results accrue: the residual gas is driven out of the matrix, leaving it in the R_0 state, and a powerful electrolytic conductivity is developed.

It will be clear from the above remarks that we have means available for isolating residual-gas actions from those resulting from electrolytic dissociation. Thus, Section 2 is concerned with valves having both active nickel and platinum cores and processed according to Table 1; these exhibit only the effects of residual gas. Section 3 is concerned with valves having platinum cores exclusively; these require additional thermal sintering to develop an adequate level of electrolytic conductivity and, incidentally, to free them from residual gas.

(2) RESIDUAL-GAS EFFECTS

(2.1) General

Except where otherwise stated, all the work reported here was carried out at the conventional matrix temperature of $1\,020^\circ$ K. Apart from variations of core metal, the S-type assemblies used all conformed to the specification set out in Section 1.1 of the present Part and were vacuum processed according to the schedule detailed in Part 1. Results are presented for individual valves, but these can be regarded as typical samples from batch results. The characteristics generally employed are the variations of diode resistance, R_d, with matrix current, I_A, and with time (with and without current load I_A). Both characteristics involve the passing of current through the matrix, but S-valves processed according to the schedule detailed in Part 1 possess negligible electrolytic conductivity—an assertion which will be justified in Section 3. The effects noted in the present Section are therefore due wholly to residual gas lying in the vacuous pores of the matrix. It is probable that this gas is carbon monoxide.

(2.2) Influence of Current Load on R_d/Time Characteristics of Active-Nickel-Cored Assemblies

Two assemblies with active nickel cores were set up for an extended time run at $1\,020^\circ$ K. In the first sample the matrix current, I_A, was maintained at 50 mA throughout the run by continuous adjustment of the potential, V_A, maintained between the core faces. The second sample was run with zero current through

* An extract of the thermal details from Table 1 of Part 1 is set out below for convenience of reference.

Approximate temperature of cores, deg K	Time, min
740	1
910	1
1 010	1
1 150	1
1 280–1 300	3–5

the matrix, apart from the short periods necessary from time to time for 'spot' measurements of R_d. The two characteristics of R_d/time are set out in Fig. 1(a) for a test run of 600 hours, and give rise to the following observations:

(a) At zero time both samples show high values of R_d.
(b) After about 400 hours both samples have fallen to the level of the characteristic resistance R_0.
(c) The rate of approach to the R_0 level is much greater in the current-loaded samples than in the unloaded sample.

The comparative behaviour of the valves is explained in the following terms. At zero time the resistance of both valves is
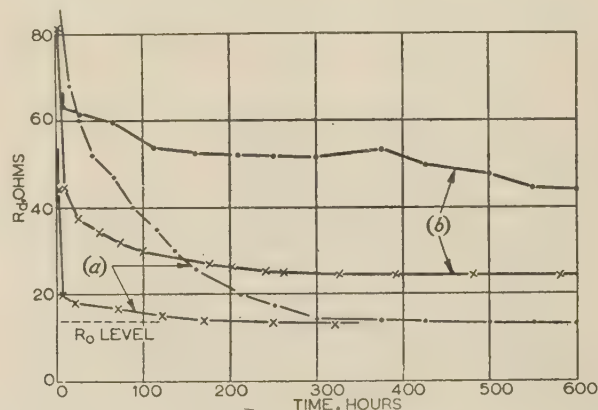
Fig. 1.—Characteristics of R_d/time for S-assemblies.

—×—×— Current load of 50 mA.
—·—·— No current load.
(a) Active nickel core.
(b) Platinum core.

high, owing to the presence of negative ions of residual gas in the matrix pores. These ions may be visualized as being adsorbed on the pore wall surfaces but free to move at random by virtue of their thermal energies, or in a preferred direction in an electric field. After some lengthy period of operating time both valves achieve the characteristic state R_0 and may then be regarded as having eliminated, by ejection or destruction, the negative ions from the matrix pores. There seem to be only three possible ways in which this elimination can occur, namely

(d) The ions can travel to the thin exposed edge of the rectangular matrix slab and escape into the surrounding vacuum (vacuum escape).
(e) The ions can move into contact with the nickel core faces and be eliminated by an irreversible chemical action (core oxidation).
(f) The ions can seek out and combine irreversibly with excess barium ions lying on the pore walls of the activated matrix (excess-barium oxidation).

In the unloaded sample in Fig. 1(a) the ions are subjected to no electric field and must therefore take up a uniform distribution throughout the matrix pores. Owing to random thermal motion, an ion may at any time strike an excess barium ion on a pore wall, collide with a nickel core face, or pass to the thin edge of the matrix exposed to the outer vacuum. The slow wastage of ions in the unloaded sample may thus be due to one or all of the three loss mechanisms. If a potential V_A is now applied across the matrix, the ions start their orderly progression towards the anodic core, and each individual ion must in turn come into intimate contact with the hot anodic nickel face. The possibility of ion elimination by core oxidation is thus enormously enhanced.

The single conclusion reached from this first experiment is that core oxidation is a powerful mechanism for residual-gas ion elimination.

(2.3) Influence of Current Load on R_d/Time Characteristic of Platinum-Cored Assemblies

The second experiment was a straightforward repetition of the first, with the exception that the core pieces were changed from active nickel to pure platinum. The characteristics for the load and no-load cases are set out in Fig. 1(b) and lead to the following observations:

(a) Neither case approaches the R_0 level in the 600-hour test period.
(b) Both cases show a decrease of R_d with operation time, but the current-loaded sample has much the higher rate in the first 100 hours.

In explaining these results the first point to be remembered is that use of platinum-cored assemblies excludes core oxidation as a mechanism of residual-gas ion elimination. Ion wastage must now depend on escape into the surrounding vacuum and excess barium oxidation.

From a comparison of the two no-load cases of Fig. 1 the following firm conclusions are drawn:

(c) The primary reason for the approach of the active nickel assembly to the R_0 state is core oxidation.
(d) The two mechanisms of vacuum escape and excess barium oxidation make a definite contribution to ion elimination, but at a very much lower rate than the mechanism of active nickel core oxidation.

The effect of current load on the platinum case is explained in much the same terms as that for the active nickel case. The negative ions are driven by the applied electric field from the general body of the matrix and are stored in a thin layer adjacent to the anodic core. The bulk of the matrix is therefore free from residual gas and is correspondingly low in resistance. The thin ion-impregnated layer of the anodic core is of high resistance and accounts for the inability of the diode to approach the R_0 state. The essential difference between the two cases is one of reversibility: in the active-nickel core assembly the ions are permanently removed from the system by chemical action on impact with the anodic core, whereas in the platinum case the ions are merely held against the core so long as an electric field exists. On removal of the field from the platinum assembly the whole of the ions can be expected to flow back into the general body of the matrix and increase its resistance to the level previously shown under zero load.

(2.4) Reversibility of R_d/Time Characteristic in a Platinum Assembly under Conditions of Load and No-Load

The starting-point of the third experiment was the platinum-cored assembly running under zero-load condition [Fig. 1(b)]. This valve continued on no-load test up to 840 hours, when its resistance had fallen to 41·6 ohms. At this stage a continuous current of 50 mA was passed through the matrix for 50 min, after which the valve was again returned to the no-load condition with occasional 'spot' measurements of resistance. After 40 hours R_d had returned to 41·8 ohms; i.e within the limits of experimental error the resistance had executed a complete cycle. Results are set out in Fig. 2 and lead to the following observations.

(a) On application of 50 mA of matrix current R_d falls to a low steady value R_{d2}, which is still, however, above the R_0 state.
(b) On reversion to the no-load condition the resistance rises from R_{d2} to re-establish itself ultimately at the original value R_{d1}.
(c) Equilibrium is established rapidly under load condition and slowly under no-load condition.

The interpretation of these observations is straightforward. Under electric stress the negative ions are pressed into a thin layer of matrix adjacent to the anodic core, leaving the bulk of the matrix gas-free and therefore low in resistance. The resistance of the thin gas-filled layer accounts for the excess of R_{d2} over R_0. No chemical actions take place at the platinum core face, and on removal of the stress the ions flow back into the
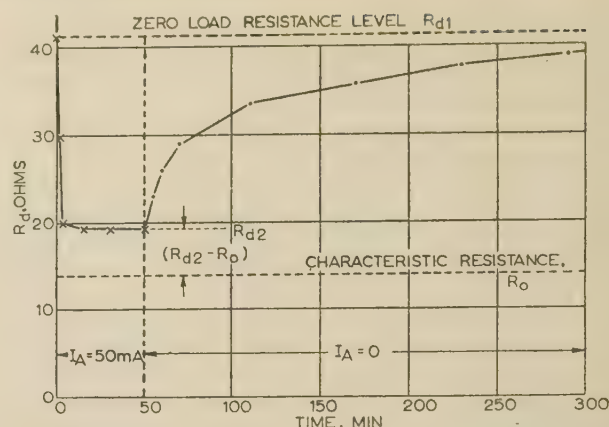
Fig. 2.—Reversibility of R_d/time characteristic for platinum-cored S-assembly.

matrix to re-establish their original uniform concentration. Under electric stress the ion movement is rapid and equilibrium is established in a few minutes; with no electric stress the return movement is limited by thermal diffusion under a concentration gradient and is therefore slow.

An obvious point of interest arising from Fig. 2 is the quantity $(R_{d2} - R_0)$, which must be some function of the matrix current I_A or the electric force applied to the matrix.

(2.5) Characteristic of R_d/I_A in the presence of Residual Gas Ions in a Platinum Assembly

In the fourth experiment a platinum-cored assembly was set up at 1 020° K and run with a current load of 20 mA until the resistance was stable with time. A characteristic of R_d/I_A was then taken, with precautions to ensure that the valve reached equilibrium at each current setting. Results are set out in Fig. 3, which includes three different samples to give an impression of
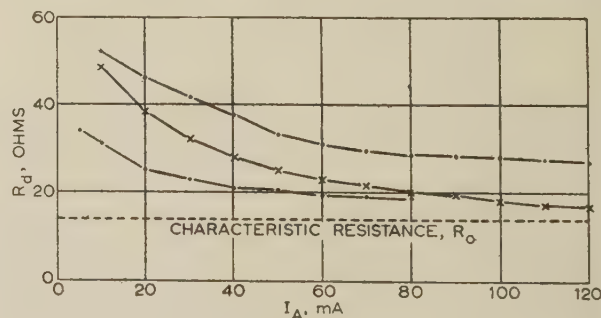
Fig. 3.—Characteristic of R_d/I_A in presence of residual gas.
Three different samples are shown.

typical experimental spread. The characteristics show that R_d diminishes as I_A increases, and may in certain cases approach the limiting value R_0.

The result will be explained in terms of the following assumptions:

(a) A standard S-type assembly at 1 020° K and free from residual gas shows the characteristic resistance R_0 of 14–18 ohms irrespective of the core material.
(b) Any departure from R_0 is due to residual gas in negative ion form.
(c) Residual gas in molecular or non-ionized form has no influence on the matrix resistance.
(d) After an ion has discharged at the anodic core surface it re-enters the matrix as an uncharged molecule, and has a finite life period as such before it becomes re-ionized. During this life period the molecule moves at random in the matrix pores by virtue of its thermal energy.

Consider a platinum-cored system containing residual gas and in a condition of equilibrium after zero-load running, i.e. the gas negatively ionized and uniformly distributed. If a current I_A is now applied, each ion moves towards the anodic core, is discharged on arrival and moves away again to begin its life period as an uncharged molecule. The R_d/I_A characteristic is supposed to depend on the magnitude of the uncharged life period, and two cases will be considered. If the life period is vanishingly small the molecule will become re-ionized as soon as it leaves the core surface, when the electric force will return it again to the core face. A cyclic state of charge and discharge will thus set in and never permit the particles to leave the immediate vicinity of the core face. Such an action should lead to a constant value of R_d near the characteristic level R_0 for all values of I_A greater than zero—a conclusion which is clearly inconsistent with Fig. 3.

Suppose now that the molecular life period, t_1, is relatively great and that a molecule can travel deeply into the body of the matrix before it becomes re-ionized. After re-ionization the ion travels back to the anodic core in time t_2, when it discharges to start on another cycle. The molecular life period, t_1, can be regarded as constant for any particular matrix temperature, but the ionic travel period, t_2, is essentially a function of the applied electric force V_A. The ratio t_1/t_2 can thus be made progressively greater as the electric force increases, and in the limit the residual gas can be made essentially molecular in character. On the assumption that the molecular form has no influence on R_d, the effect of increasing I_A is therefore to decrease the negative-ion content of the matrix until its resistance approaches the characteristic state R_0.

(2.6) Influence of Thermal Treatment on Residual-Gas Content of Platinum-Cored Assembly

So far, it has proved impossible to bring a platinum-cored assembly permanently into the state of the characteristic resistance R_0. The state can indeed be approached in certain cases by applying a sufficiently great electric force to move the bulk of residual gas ions onto the passive anode surface, but on release of the force the ions flow back into the matrix, which resumes its former high-resistance state. The picture arising is that of a mass of gas capable of rapid movements throughout the porous matrix but hardly able to escape from it or to combine chemically with it. The object of the fifth experiment was to examine this picture in greater detail.

A standard S-type assembly with platinum cores was set up at $1\,020°\,K$ and its resistance was measured. The matrix temperature was then raised to $1\,300°\,K$ for 1 min, after which it was lowered again to $1\,020°\,K$ for remeasurement of resistance. This cyclic sequence of high temperature treatment and normal temperature resistance measurement was continued until the resistance approached a constant value. The experiment was repeated on additional assemblies for a range of temperatures, and two typical results are set out in Fig. 4, where the abscissa records total time of high-temperature treatment. It is apparent from the characteristics that such treatment brings the platinum-cored assembly into the characteristic state R_0 and that the approach to this state is more rapid with higher temperatures. The platinum assembly has thus the same basic characteristic as the active nickel assembly. On the assumption that achievement of the characteristic state is due to the elimination of residual gas from the matrix pores, it can be expected that the degassed platinum system will retain its resistance R_0 for an indefinite period of operation at $1\,020°\,K$. Fig. 5 records a typical result of a platinum system degassed at $1\,450°\,K$ and running under zero load with occasional 'spot' measurements of resistance; the characteristic is recorded over 600 hours in the illustration, but
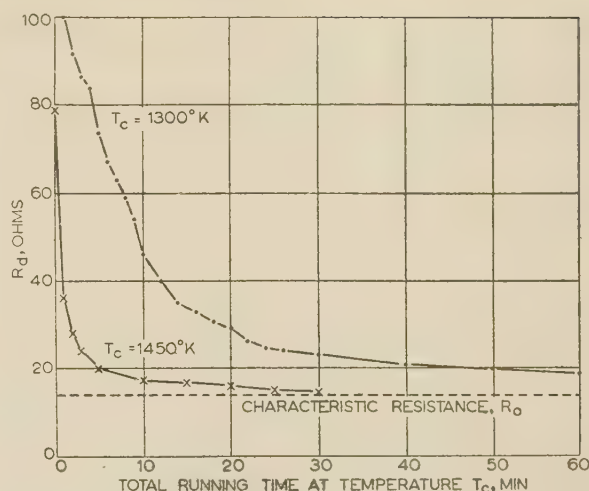
Fig. 4.—Influence of thermal treatment on residual gas content of platinum-cored assembly.
The resistances are measured at $1\,020°\,K$.
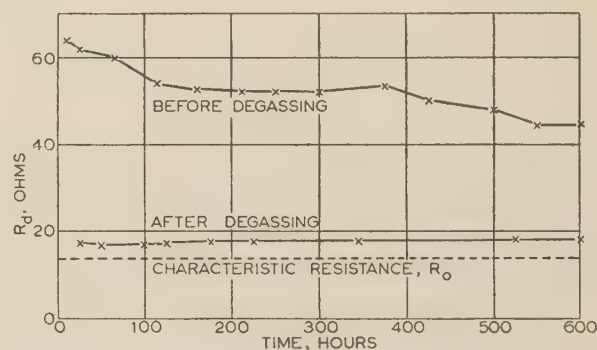
Fig. 5.—Characteristic of R_d/time under zero load, before and after degassing at $1\,450°\,K$.

tests up to $2\,000$ hours show no significant change. Behaviour is essentially similar to that of an active nickel system.

Brief consideration will now be given to the mechanism of gas elimination which has led to the achievement of the R_0 state in the platinum system. In the absence of core oxidation, gas must be withdrawn from the matrix pores by escape into the surrounding vacuum and/or by chemical combination with a fraction of the excess barium metal of the activated matrix itself. If the vacuum escape mechanism is involved, it should be possible to detect the escaping gas by some form of detector enclosed within the vacuum envelope of the S-type valve. Experiments on such lines have been carried out using a conventional form of oxide-cathode diode as detector and relying on the 'poisoning' action of the gas to change its space-charge-limited current, I_d. The diode detector is processed simultaneously with the S-type assembly and both systems are left to stabilize for a few hours at $1\,020°\,K$. The detector diode is run with a space-charge-limited current and the S-type assembly measured for R_d. The temperature of the S-assembly is then raised from $1\,020°$ to $1\,350°\,K$ for 1 min, after which it is returned to $1\,020°\,K$ for remeasurement of R_d. Two concurrent effects are observed: R_d has fallen and a powerful 'poisoning' action on the detector diode has reduced I_d to a small fraction of its initial steady value. After thermal reactivation of the detector diode the sequence is successively repeated until the resistance of the S-assembly has fallen to R_0, when the 'poisoning' action on the detector diode ceases to influence I_d.

The following conclusions are drawn from such experiments:

(a) Reduction of resistance to R_0 by thermal action is accompanied by a powerful evolution of residual gas.

(b) The quantity of gas that can be driven from the matrix is limited, and the achievement of the characteristic level R_0 coincides with the approach to gas exhaustion.

(c) Once the residual gas has been driven out of the matrix into the surrounding vacuum it does not re-enter to any appreciable extent (cf. Fig. 5).

(2.7) Residual-Gas Action in an Alternating Electric Field

A platinum system with a matrix containing residual gas was used for the investigation in which R_d is explored in a 50 c/s electric field. The characteristic examined is that of R_d/I_{AC}, where I_{AC} is the alternating current traversing the matrix. There are clearly two states of ionic distribution under which the characteristic can be measured—one with uniform concentration of ions throughout the matrix and the other with all the ions compressed into a thin layer adjacent to the anodic core. Both states will be examined, for they lead to significantly different forms of R_d/I_{AC}. Fig. 6 sets out a representative example which was handled on the following lines:

(a) Valve No. S–111 was processed to the standard schedule and run under zero load until 'spot' resistance measurements at 1 020° K showed that it was steady at $R_1 = 50$ ohms. This zero-load condition leaves the negative ions of residual gas uniformly distributed throughout the matrix.

(b) A characteristic of R_d/I_{AC} was taken and recorded as phase A in Fig. 6, showing a remarkable increase in resistance under a.c. loading. On removal of the load the resistance returned almost immediately to the zero-load value R_1.

(c) The valve was then conditioned under d.c. loading, and phase B records the R_d/I_A characteristic up to 100 mA of direct current. This operation forces the residual ions against the anodic core surface, leaving the general body of the matrix gas-free and close to the characteristic level R_0.

(d) The valve was next changed over to the a.c. condition and a characteristic of R_d/I_{AC} taken up to 100 mA. The operation is recorded as phase C and shows R_d to be virtually independent of I_{AC}—a radically different result from the a.c. case of phase A.

(e) The valve was finally run under zero current load (operation X) for 20 hours to re-establish R_1. The application of a.c. loading then results in the abnormal resistance rise shown in phase D.

The pattern of behaviour under 50 c/s loading is thus shown to be systematic. With the ions removed from the body of the

Fig. 6.—Characteristic of R_d/time under different ion distributions.

matrix and forced on to the anodic core surface, the resistance is largely independent of current and is near the R_0 level; with a uniform distribution of ions throughout the matrix, a drastic and abnormal increase of resistance is observed.

A rather different approach to the phenomenon is shown in Fig. 7 for valve No. S–114. This valve was run under a d.c. load of 20 mA and showed a constant resistance of about 30 ohms
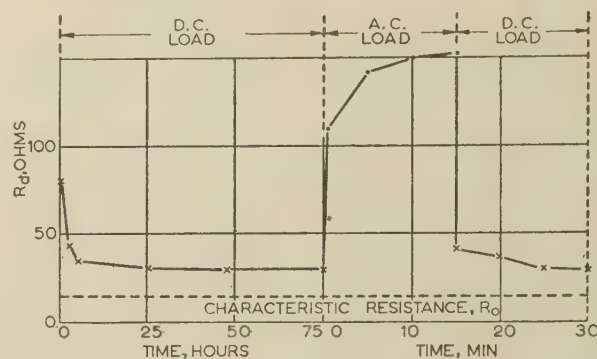
Fig. 7.—Characteristic of R_d/time under d.c. and a.c. loading.

over a period of 50 hours. The resistance is, of course, well above the R_0 level, and shows that the general body of the matrix still retains gas ions within its pores. At this stage the d.c. load was replaced by one of 50 mA at 50 c/s and resulted in a rapid rise of resistance to 150 ohms, almost the whole of the increase occurring in the first few minutes. On reverting to the d.c. condition the resistance collapsed back to its former steady value of 30 ohms.

The abnormal rise of a.c. resistance in the case of distributed ions is explained in terms of Fig. 8, which shows schematically
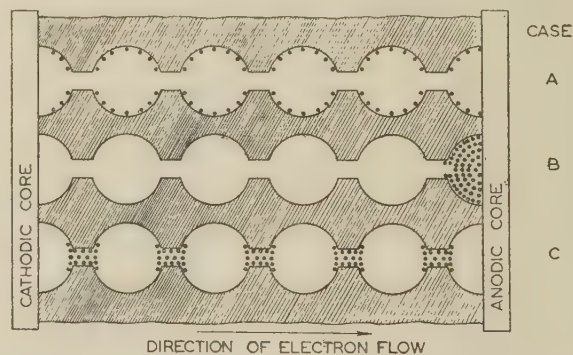
Fig. 8.—Residual gas ion distribution in the matrix pores.

an imaginary cross-section through the matrix. The pore system is visualized as a series of caverns linked by narrow tunnels to form a chain of vacuum passages from the cathodic to the anodic core. Almost the whole of the electron current I_A is assumed to pass through these passages, which also accommodate the free-moving negative ions of residual gas. The ions of gas are shown as black circles, and case A in the Figure shows the uniform distribution following a period of operation under zero current load. The d.c. resistance in case A will be greater than the characteristic level R_0 by an amount dependent on the ion concentration. Case B shows the ions pressed up against the anodic core under an adequate electric force and resulting in an effectively gas-free matrix of resistance R_0. Suppose now that a 50 c/s field is applied to the condition of case B: the result is simply an oscillation of the ions in the immediate vicinity of the anodic core, leaving the matrix itself gas-free and at the R_0 level. This a.c. condition will, of course, last only for a limited time, for the ions will tend to diffuse back into the matrix under their own thermal energies, which exist independently of the
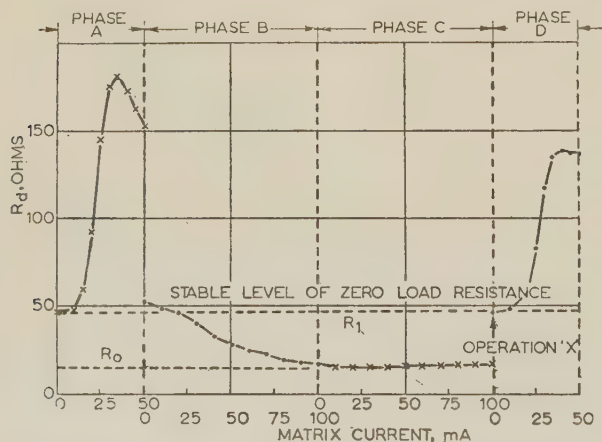
applied electric force. The measurements of phase C in Fig. 6 were, for example, carried out within a period of 2–3 min, but even in this limited time there are signs of back-diffusion of ions.

Under the condition of uniform ion distribution of case A the following alternating-current action is supposed to occur. During a positive half-cycle a negative ion, moving easily and without obstruction within a cavern, approaches the entrance to the outlet tunnel. Once within the narrow tunnel, however, the ion suffers repeated collisions with the tunnel walls and its rate of progression is drastically reduced. The effect of this action is to concentrate the bulk of the ion population of the matrix within the tunnels, where it can exert its maximum influence on the main electron stream, I_{AC}. On reversal of the electric field the ions leave the outlet tunnels, traverse the caverns at speed, and start electrostatically blocking the inlet tunnels. The action is shown schematically as case C in Fig. 8.

Two points of interest arise from the a.c. picture: the influence of frequency and of the magnitude of the current I_{AC}. At very low frequencies an ion is able to pass through every cavern and every tunnel during each half-cycle and we have the same effect as in the slow d.c. reversal examined in Part 1 of the paper, i.e. each reversal sees the bulk of the ions move into the vicinity of the anodic core. At a sufficiently high frequency it seems clear that the heavy negative ion will be unable to travel the length of a cavern during a half-cycle, and the a.c. effect should therefore become progressively less as frequency increases.

The influence of alternating voltage swing is simple to predict. With a small electric force the negative ion is unable to travel the full length of a cavern and only ions in the immediate vicinity can enter a tunnel mouth. As the electric force is increased, more ions are able to make the cavern crossing until all are so capable and a saturation effect should occur.

The two suggested effects have been given brief experimental attention. A standard S-type assembly with platinum cores was set up at $1\,020°\,K$ and shown to have a resistance of about 40 ohms running under zero current load. A constant alternating current of 25 mA was then passed through the device and resistance measurements were made at frequencies ranging from 50 to 5 000 c/s. The characteristic (Fig. 9) indicates a falling resistance with increase of frequency, and is therefore consistent with

The effect of voltage or current swing at a constant frequency of 50 c/s is examined in Fig. 10, which shows the a.c. resistance as a function of current amplitude. The predicted saturation
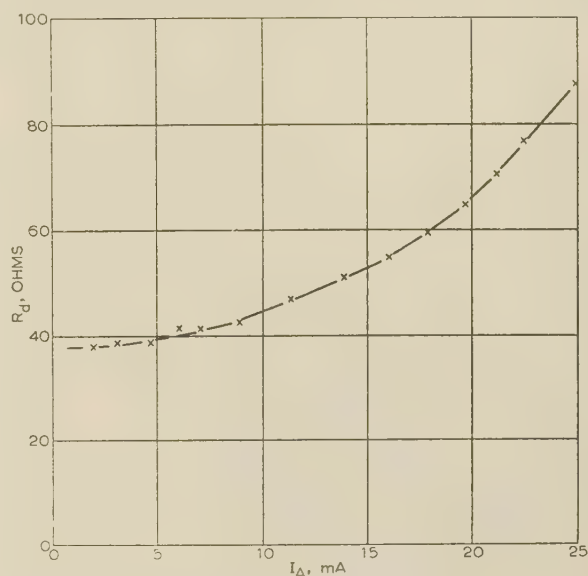
Fig. 10.—Characteristic of R_d/I_A under a.c. loading at 50 c/s.

effect as current swing increases has not been investigated, for it is complicated by temperature considerations arising from power dissipation within the matrix. The two resistance maxima shown in the R_d/I_A characteristics of Fig. 6 are, for example, due primarily to a progressive rise in matrix temperature rather than to the onset of a saturation effect.

(2.8) Electron-Emission Density Gradient

The passage of current through an S-type assembly with active nickel cores results in a resistance gradient across the matrix if residual gas is present. It is proposed now to show that this positive resistance gradient may coincide with a negative gradient of emission density across the matrix.

Fig. 9.—Frequency dependence of R_d at constant current.
$I_A = 25\,\text{mA, a.c.}$

the tunnel-blocking hypothesis. Any frequency effect due to the natural capacitance of the system can be disregarded, since it is of the order of a megohm at 5 000 c/s.
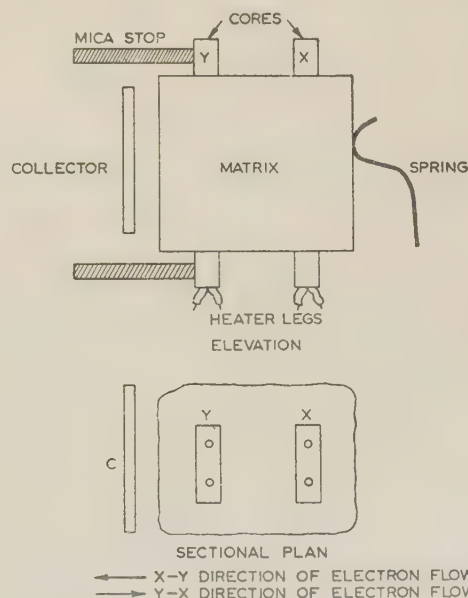
Fig. 11.—Schematic of S-valve with external collector electrode.

An S-assembly of standard dimensions is shown in Fig. 11 modified to include a nickel plate collector C placed in vacuo close to core Y. The (BaSr)O matrix has been sprayed on all sides of the two active nickel cores X and Y, and it is therefore possible to explore the total or temperature-limited emission from the outer side of core Y to the collector C. Furthermore, this side operation can be conducted with the main electron flow I_A between the cores in either the X–Y direction or the Y–X direction.

Suppose that an electron flow I_A occurs between the two cores in the X–Y direction. Negative ions of residual gas will congregate in the vicinity of the Y-core and are presumed to depress the emission density of the trailing half of the matrix. The build-up of negative ions on the inner face of the Y-core will furthermore lead to a flow through the matrix pores to the outer face of Y, where the effect should be detected by a falling value of total emission I_T measured between Y and C. If the main matrix current is reversed the ions should leave the environment of Y and congregate around the X-core, resulting in a recovery of the total emission I_T as the operation proceeds.

A typical result is shown in Fig. 12, where a constant matrix electron flow, I_A, of 50 mA at 1 020° K has been passed in the Y–X direction for 7 hours. During this period 'spot' tests of total emission from Y to C were made by momentarily lowering the system temperature to 700° K and measuring I_T to electrode C under a saturating potential of 10 volts. At the end of this first 7-hour run the current of 50 mA was reversed into the X–Y direction for 7 hours and the 'spot' tests of I_T were again taken at suitable intervals of time. This operation was continued until two complete cycles were covered.
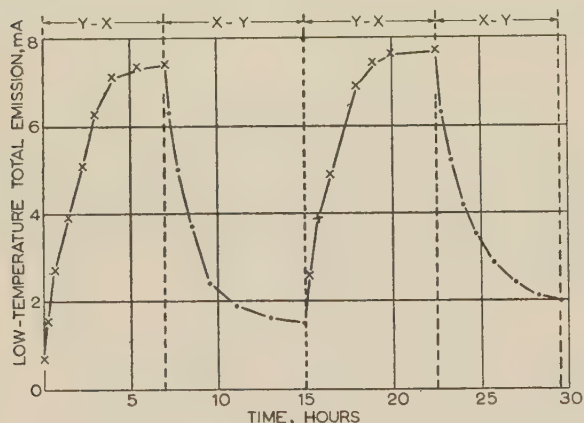
Fig. 12.—Local variation of thermionic activity due to change in direction of electron flow.
Low-temperature total emission measured at 700° K.

The result in Fig. 12 shows that the total emission that can be drawn from the vacuum surface of the trailing section of the matrix is much smaller than that available from the leading section. It seems reasonable to draw the following conclusions:

(a) The internal emission density within the pores is high in the leading section and low in the trailing section.

(b) An emission density gradient occurs within the matrix and coincides with the conductivity gradient.

That the phenomenon is due to the presence of residual gas ions is proved experimentally by its disappearance when the active nickel-cored system shown in Fig. 11 has run for a sufficient length of time to achieve the state of characteristic resistance R_0.

(2.9) Summary

(a) Residual-gas effects studied are probably due to the presence of carbon monoxide in the matrix pores.

(b) The gas molecules become negatively ionized on entering the pores.

(c) The ions are capable of rapid thermal motion at 1 020° K throughout the pore system.

(d) The ions are highly mobile in an electric field.

(e) The ions are chemically stable within the activated matrix pores.

(f) The ions are chemically stable in the presence of a platinum core at 1 020° K but are destroyed by an active nickel core.

(g) The resistance to electron flow of a unit cube of matrix increases as its concentration of residual gas ions increases.

(h) The electron density within the vacuous pores of a unit cube of matrix decreases as its concentration of residual gas increases.

The resistance of a gas-free matrix is R_0, and any excess value is due to the presence of residual-gas ions, which arrange themselves in conformity with an applied electric field. The actual physical mechanism whereby the negative ion causes a local increase of resistance may be either a direct electrostatic repulsive action on the electron stream or a local suppression of electron emission from the pore walls. An attempt will be made to distinguish between the two processes in a later Part of the paper.

(3) ELECTROLYTIC GAS EFFECTS

(3.1) Introduction

The passage of electrons through an activated oxide matrix may in certain circumstances be accompanied by the dissociation of a fraction of the matrix into its component ions. Consider the case of a platinum-cored S-type assembly of standard dimensions which has been cleared of residual gas by thermal action. The device will be in the characteristic state R_0 and at 1 020° K will have a resistance of about 16 ohms, with 8 ohms in the leading section and 8 ohms in the trailing section. Suppose now that a current-dependent matrix dissociation occurs, with barium metal congregating in the leading and oxygen in the trailing section. The electron density in the pores of the leading section will increase and its resistance will fall; in the trailing section the electron density will be depressed by the oxygen and the resistance will increase. If the resistance effects in the two sections are equal in magnitude, the overall resistance of the device will remain unchanged at the R_0 level. Experiment will show, however, that the section resistances do not change by anything like equal amounts: further activation of the leading section has little effect on its resistance, whereas oxygen deactivation of the trailing section results in a relatively large increase of resistance ranging up to 100 ohms or more. Electrolytic action may therefore be expected to cause an increase of overall resistance of the device, and the magnitude of the increase will probably depend both on current magnitude and time. Two prerequisites are essential to a practical demonstration of the action, namely that the matrix should possess an adequate measure of electrolytic conductivity, and the anodic core must be such that the electrolytic oxygen in the trailing section is not removed chemically as fast as it is produced.

The S-valves processed according to the standard schedule of Table 1 of Part 1 are fully activated in the sense that a current density of 2–6 amp/cm^2 can be drawn from the exposed oxide surface at 1 020° K. Despite this high level of thermionic activity, however, such valves have a negligible electrolytic conductivity and the phenomenon becomes apparent only after the matrix has aged for some hundreds of hours at 1 020° K. This slow growth of the conductivity is thought to be due to a sintering action in the matrix during which chains of contiguous oxide particles become welded together to give a more effective medium for the solid transport of ions. The sintering action can, of course, be hastened to any desired extent by increasing the

maximum temperature to which the matrix is subjected during processing. If, for example, the standard processing temperatures set out in Table 1 are increased by 100° K, the tubes will enter service with no sensible change of thermionic activity but with greatly enhanced electrolytic conductivity. It is by such means that the platinum-cored S-valves discussed in the present Section were prepared.

(3.2) Development of Electrolytic Action

During the standard processing schedule of Table 1 an S-valve is held for 3–5 min at a maximum temperature of 1 300° K. Such a valve fitted with platinum cores was prepared and its characteristic of R_d/I_A measured at 1 020° K over a current range of 10–120 mA. The tube was then subjected to a further period of 15 min of sintering at 1 300° K under zero load and was remeasured for R_d/I_A characteristic at 1 020° K. This procedure was continued until the R_d/I_A characteristics following periods of 15, 30, 60 and 120 min of sintering had been measured. The results are set out in Fig. 13, where the figure attached to each
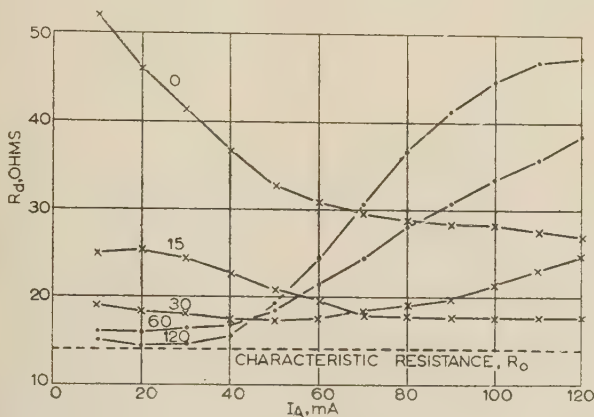
Fig. 13.—Characteristics of R_d/I_A showing onset of electrolysis with matrix sintering time.

The figure on each curve gives the total sintering time, in minutes, at 1 300° K.

characteristic denotes the total sintering time at 1 300° K in excess of the time associated with the initial processing. The following observations arise:

(a) With zero additional sintering time the resistance R_d falls continuously with increase of current—a typical residual-gas action in the absence of electrolysis.

(b) After 15 min sintering much of the residual gas has been thermally ejected and the resistance has approached closer to the R_0 level.

(c) After 30 min sintering R_d approaches R_0 at low values of I_A, but above 50 mA the characteristic changes in slope and R_d begins to increase with increase of I_A; this change indicates the onset of electrolysis.

(d) After 60 and 120 min sintering the residual-gas effect becomes progressively smaller and the rise of R_d with I_A progressively greater.

The sequence of characteristics show a smooth transition from a matrix containing residual gas but possessing no electrolytic conductivity to one which is free from residual gas but is powerfully affected by an accumulation of electrolytic oxygen in its trailing section. It is a fortunate experimental circumstance that the sintering of the matrix to develop electrolytic conductivity has the simultaneous effect of freeing the matrix from residual gas. Separation of the two gas actions is therefore simple, and they might be defined in terms of slope of the two R_d/I_A characteristics, thus:

Residual-gas action: R_d/I_A characteristic has negative slope.
Electrolytic gas action: R_d/I_A characteristic has positive slope.

Although a platinum-core system is used to demonstrate electrolysis, the cores in themselves have, of course, nothing to do with the electrolytic action: by rejecting oxygen they merely enable the action to be detected in terms of resistance rise. The actions, can, in fact, be detected in an active nickel-core system if the experiment is conducted at a sufficiently low temperature to inhibit the absorption of electrolytic oxygen by the anodic core.

(3.3) The Time Function of R_d at Constant Current

A standard S-valve with platinum cores was prepared in the usual way and sintered at 1 450° K for 15 min. This operation leaves the valve with a resistance close to the R_0 value and with a high level of electrolytic conductivity. The following sequence of measurements was then undertaken. With zero current I_A the two core-heater voltages were set to give a uniform temperature of 1 020° K across the matrix, and these heater voltages were maintained constant throughout the experiment. A constant current of 20 mA was next applied to the matrix for 50 min and readings of R_d were taken at intervals. At the end of the time run the current was reduced to zero and the temperature of the matrix was raised for 1 min to 1 300° K to bring the resistance back to the R_0 level. The experiment was then repeated at a constant current of 30 mA for 50 min, and so on until an adequate range of current settings had been covered. Results are set out in Fig. 14(a) and one case extended in time to 250 min is shown in Fig. 14(b). The following conclusions are drawn:

(a) At any particular time R_d is a function of I_A.
(b) R_d is a function of time at constant I_A.
(c) After a sufficiently long period of operation at constant current R_d approaches a constant value, i.e. it ceases to be a function of time.

Once the resistance has reached its high level it seems to stay there indefinitely: at least, over a test period of 2 000 hours there is no significant trend.

The results are explained in the following terms. On application of a constant voltage across the matrix an electrolytic flow of ions is set up, and positive ions of barium and negative ions of oxygen begin to discharge at constant rate at the cathodic and anodic cores respectively. Initially we will suppose that after discharge the two elements remain in atomic form, i.e. they will move under the influence of a concentration gradient but not an electric field. As electrolysis proceeds a concentration gradient of oxygen atoms builds up off the anodic core surface and a gradient of barium atoms off the cathodic core surface. Under the influence of the gradients the two sorts of atom diffuse forward into the body of the matrix, where they meet to recombine as barium-oxide molecules. Diffusion and recombination therefore set a limit to the atomic concentrations at either core face, and lead ultimately to an equilibrium condition wherein the rate of recombination is equal to that of electrolytic dissociation. Under this equilibrium state there is no change with time in the total number of barium oxide molecules in the matrix, and the atomic concentrations at the two core faces are at the highest possible level.

So far, only atomic movements and concentrations have been considered, and it is postulated in Section 2.5 that these have no influence on the resistance of the matrix: only negative ions can cause a resistance rise. It has been shown, however, that free oxygen within the pore system has a cyclic life in alternate atomic and negative-ion forms, and that the relative life periods are constant in a constant electric field. A concentration of oxygen atoms in the matrix therefore implies a proportionate concentration of negative ions, and any time changes of an oxygen concentration are thus followed by similar time changes of R_d. On application of a constant electric force to the matrix, R_d should therefore show a steady increase at an ever-decreasing rate up to some saturation level, at which it should remain indefinitely.

(a)

(b)

Fig. 14.—Time function of R_d at constant current.

Fig. 15.—Electrolysis and recombination.

At 150 hours the current is increased from zero to 50 mA for $1\frac{1}{2}$ hours, and then decreased to zero again.

(3.5) Influence of Temperature on R_d

So far, no mention has been made of the influence of matrix temperature on the equilibrium condition described in Section 3.3, and it is proposed now to give the matter brief consideration. We will suppose that the electrolytic conductivity is invariant with temperature in the working range 950–1 100° K. (This assumption is obviously wrong, but its use as an analytic tool will be justified later in the argument.) If a constant direct voltage, V_A, is impressed across the matrix, ion separation occurs at a constant rate independent of all other considerations, including a temperature swing from T_1 to T_2. Once the equilibrium condition has been established, the electrolytic dissociation rate is exactly balanced by the recombination rate, which is fixed only by the applied potential V_A. To meet the equilibrium condition it is likewise clear that the flow rates of both sorts of atom from their respective electrode sources to the common recombination sink must also be equal, to each other and to the dissociation and recombination rates. Everything in the equilibrium system is therefore fixed by V_A with one exception—the concentration gradients of the two sorts of atom. These gradients are related to the fixed flow rate by flow impedances, and it will be apparent that the flow rate will be unaltered by wide variations in the concentration gradients, provided that these are accompanied by proportionate changes in the flow impedances. The basis of the flow impedance is the mean thermal energy of the migrating atoms, and the impedance will therefore fall as the system temperature rises. Suppose now that a system, in equilibrium at temperature T_1 and potential V_A, has its temperature increased to T_2. The mean thermal energy of the atoms will increase and the flow impedances will decrease, and in order to maintain the fixed flow rate the concentration gradients will automatically fall in exact proportion. The fall in the gradients will furthermore lead to a fall in the concentrations of the two sorts of atom and an equivalent rise in the total number of barium-oxide molecules. At T_2 there will therefore be few oxygen atoms in the matrix and likewise fewer negative ions. An increase in temperature will therefore result in a decrease in the equilibrium value of R_d.

We will return now to discuss briefly the earlier assumption of invariancy of electrolytic conductivity with temperature. If, in passing from a lower temperature T_1 to a higher temperature T_2, the conductivity increases at a certain rate, will this increase invalidate the hypothesis developed in the previous paragraph? The answer to this question lies in the relative rate at which the conductivity (or dissociation rate) increases and the flow impedance decreases. If the positive temperature coefficient of conductivity is greater than the negative coefficient of flow impedance, R_d will increase with temperature; if the two coefficients are equal, R_d will be invariant; if the conductivity

(3.4) Recombination

The application of electric force to a sintered matrix leads to ion separation, with a consequent rise in matrix resistance. On release of the force the separated ions or atoms should diffuse together, recombine as barium oxide, and cause the resistance to return to its initial value. An experiment illustrating the action is shown in Fig. 15. A sintered S-valve with platinum cores was run for 150 hours at 1 020° K under zero load, and 'spot' tests of resistance showed that it was constant at about 18 ohms. A constant current of 50 mA was then passed through the device for 90 min, and the resistance was observed to rise steadily to 42 ohms. The current was finally reduced to zero and 'spot' measurements of resistance showed that in about 6 hours it had returned to its former steady value of 18 ohms.

One essential requirement of the above experiment is that the two sorts of ion separated by electrolysis are wholly retained in the matrix system. If the electrolytic action is prolonged to 50 hours or more, R_d will not return completely to its former value on zero-load running. This effect is due to a 'leakage' of barium metal out of the matrix and into solution in the platinum cathodic core.

coefficient is the smaller, R_d will fall with increase of temperature. Recourse to experiment will, in fact, show that R_d falls drastically with increase of temperature. The assumption that the conductivity is constant, or its rate of change negligibly small, seems then an admissible artifice in the development of the hypothesis.

Direct experimentation on the temperature dependence of R_d is a simple matter. Fig. 16 shows three characteristics of R_d/I_A
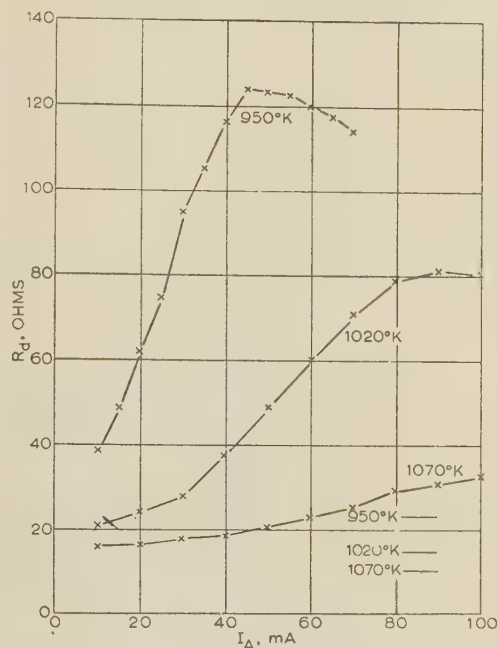
Fig. 16.—Characteristics of R_d/I_A for different matrix temperatures.

Main graph for sintered matrix: small lines in the corner show characteristics of non-sintered standard assembly, free from electrolytic dissociation.

for a sintered S-valve with platinum cores. The temperature recorded against each characteristic corresponds to the core temperature measured with zero current in the matrix. The rapid decrease of R_d with increase of temperature will be apparent. The three short lines at the bottom right-hand corner of Fig. 16 show the change of resistance over the same temperature range of a standard S-valve which is free from the effects of electrolytic dissociation. The difference in the two sets of characteristics may therefore be regarded as being due wholly to dissociation action.

The appearance of maxima in the characteristics of Fig. 16 is of no basic interest. It is due to self-heating of the system by the applied power $V_A I_A$, which causes a progressively rapid temperature rise as the current increases. The phenomenon can be avoided by adjusting the core heater power to maintain constancy of matrix temperature at each new setting of I_A.

(3.6) A.C. Phenomena

The behaviour of a standard S-assembly with platinum cores under a 50 c/s load is illustrated in a crude but effective manner in Fig. 17(a). The valve was arranged in series with a nominal 100-ohm limiting resistor across a constant 20-volt a.c. or d.c. supply. The system temperature was set at 1 050° K under zero load and thereafter committed first to a.c. and then to d.c. loading. For the first period of 360 hours under a.c. load the valve passed a fairly steady current of 140 mA, corresponding to a valve resistance of about 40 ohms. On the change to d.c. load the current fell over a period of 140 hours to 20 mA, corresponding to a valve resistance of about 1 000 ohms. Batch tests extending the a.c. load period up to 2 000 hours show
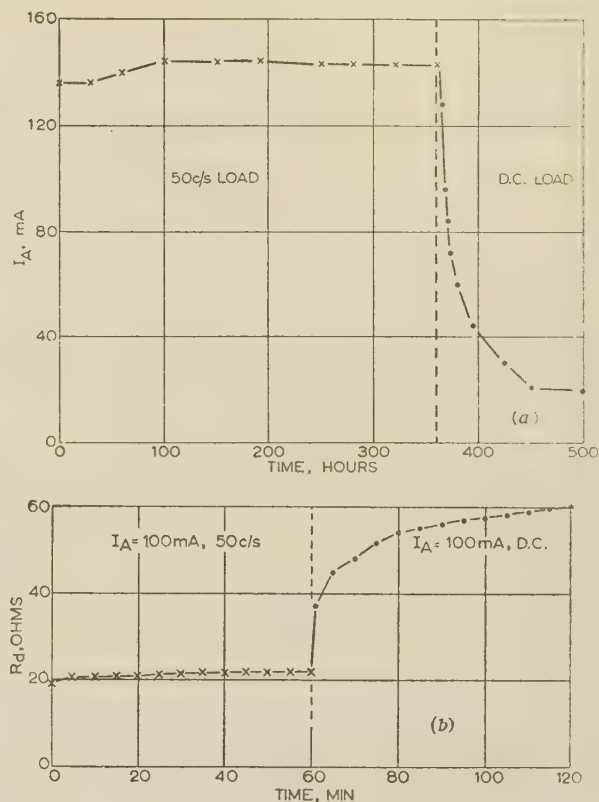
Fig. 17.—Platinum-core system under a.c. and d.c. load.

similar results, with steady resistances in the range 30–40 ohms followed by a dramatic increase in resistance on the change to d.c. loading.

A second experiment is shown in Fig. 17(b), with a standard S-valve with platinum cores which has been sintered to develop electrolytic conductivity. The zero-load matrix temperature was set at 1 020° K and the valve then put on load for two successive periods, the first of 60 min under a 50 c/s load of 100 mA and the second of 60 min under a d.c. load of 100 mA. The comparative behaviour under the two current conditions will be self-evident.

From such experiments as these it will be clear that, in absence of residual gas and in presence of an electrolytic conductivity, a 50 c/s load has negligible effect on R_d. Electrolytic dissociation is, of course, taking place, but the two sorts of ion are released in equal numbers in the same places where immediate recombination occurs, i.e. the lifetime of free oxygen in the pore system is negligible.

(3.7) Summary

Conclusions reached in the present Part are summarized in the following form:

(a) Electrolysis is present in a well-sintered BaSrO matrix under an electric force. The degree of sintering required is probably that achieved in most practical oxide cathodes.

(b) If the electrolytic oxygen is retained within the matrix system, it will cause a drastic increase in matrix resistance in a direct electric field but will have no effect in an alternating field.

(4) ACKNOWLEDGMENT

A PRECISION DUAL BRIDGE FOR THE STANDARDIZATION OF ADMITTANCE AT VERY HIGH FREQUENCIES

By D. WOODS, Associate Member.

SUMMARY

The need is stated for a more accurate means of admittance or impedance measurement in order that a general improvement can be made in the accuracy of measurement of other electrical quantities at radio frequencies. A description is given of a precision twin-T dual bridge having a range of admittance measurement associated with coaxial systems between 3 Mc/s and 300 Mc/s. The inaccuracy of measurement does not exceed $0 \cdot 2\%$ on either component of a complex admittance at a frequency of 200 Mc/s. The calibration of the bridge is based on a range of coaxial susceptance standards whose parameters are calculated from length and time measurements. The dual feature of the bridge enables the true r.f. conductance of its internal standard resistor to be established in terms of these standards; the uncertainty of this determination is not greater than $0 \cdot 1\%$ for frequencies up to 250 Mc/s.

The importance to be attached to the design of coaxial terminations for measuring instruments and standards at these frequencies is emphasized. A practical design is detailed for a coaxial terminal for precision admittance and impedance measuring instruments, and mention is made of a proposed universal coaxial connector system for precision radio-frequency measuring instruments of all types.

LIST OF SYMBOLS

C General Case. Static capacitance of a capacitor.

Particular Cases.

    (*a*) Total static capacitance of a standard coaxial capacitor due to an undistorted radial field, i.e. neglecting the fringe capacitance.

    (*b*) Capacitance per unit length of the main variable capacitors C_a and C_b.

$C' = C \tan \beta l / \beta l$ = Effective capacitance of a standard coaxial capacitor neglecting the fringe capacitance.

$$C'' = (C' + C_f) \Big/ \left(1 - \frac{C_f}{C'}\tan^2 \beta l\right) = \text{Effective capaci-}$$

    tance of a standard coaxial capacitor terminated by the fringe capacitance C_f.

$\Delta C'' = C'' - C_f$ = Change in capacitance at the bridge terminals caused by the connection of a standard coaxial capacitor terminated with the fringe capacitance.

C_a, C_b = Static capacitance of the main variable capacitors.

$\Delta C_a, \Delta C_b$ = Change in C_a and C_b, respectively.

    B-arm measurements $\Delta C_b = C_{b1} - C_{b2}$ (susceptance dial).

$$\Delta C_a = C_{a1} - C_{a2} \text{ (conductance dial).}$$

    A-arm measurements $\Delta C_a = C_{a1} - C_{a2}$ (susceptance dial).

$$\Delta C_b = C_{b2} - C_{b1} \text{ (conductance dial).}$$

$\Delta C_{aA}, \Delta C_{aB}$ = Change in C_a for A- and B-arm measurements, respectively.

$\Delta C_{bA}, \Delta C_{bB}$ = Change in C_b for A- and B-arm measurements, respectively.

C_{at}, C_{bt} = 'Set-zero' capacitors.

C_d = Discontinuity capacitance in the plane of the bridge terminals owing to the presence of the cone between the terminals and an unknown admittance.

C_f = Terminal fringe capacitance.

$C_i = C_x - C_d$ = Input capacitance of the cone or line, terminated by the unknown admittance, in the plane of the junction between the cone and the terminals and corrected for the discontinuity capacitance C_d.

$C'_n = -1/\omega^2 Z^2 C'$ = Effective capacitance of a short-circuited coaxial standard capacitor.

$\Delta C'_n = C'_n - C_f$ = Change in capacitance at the bridge terminals caused by the connection of a short-circuited coaxial standard capacitor.

C_0 = Total static capacitance of C_a or C_b at $\theta = 0$.

C_r = Capacitance of the admittance connected to the remote end of the cone or line, i.e. the capacitance of the unknown.

C_{ss} = Static shunt capacitance across G_s.

C_{sg} = Effective shunt capacitance of G_s as modified by the self-inductance L_{ss}, the earth inductance L_g and the mutual inductance L_m.

C_t = Total capacitance of the terminals for an undistorted radial field.

C_x = Capacitance of the unknown in the plane of the bridge terminals.

C_{xa}, C_{xb} = C_x related to the A or B-arm terminals.

$\delta C_x = \omega^2 L_t C_f (C_f + C_t)$ = Correction term arising from the transformation of an admittance through the terminal impedance.

C_1, C_2, C_3 = Static capacitance of the three series capacitors in the bridge.

$C_5 = 2(C_1 + C_2 + C_1 C_2/C_3 - C_3) = 5 \,\mathrm{pF}$, when $C_1 = C_2 = \tfrac{1}{2}C_3 = 5 \,\mathrm{pF}$.

g_a, g_b = Residual conductance of the shunt elements due mainly to L_a and L_b.

$G_i = G_x$ = Input conductance of the cone or line, terminated by the unknown admittance, in the plane of the junction between the cone and the bridge terminals.

G_r = Conductance of the admittance connected to the remote end of the cone or line, i.e. the conductance of the unknown.

$G_s = G_{ss}(1 + z)$ = Effective shunt conductance of the bridge resistor due to all causes.

G_{ss} = D.C. conductance of the bridge resistor.

G_x = Unknown conductance in the plane of the bridge terminals.

$G_x = G_i$ when a cone or line is used between the bridge terminals and the unknown admittance.

$G_{xa}, G_{xb} = G_x$ related to the A or B-arm terminals.

$K_1, K_2, K_3 =$ Effective capacitances of C_1, C_2 and C_3 due to their residual inductances.

L_A, L_B or $L =$ Inductance per unit length of the main variable capacitors C_a and C_b.

$L_a, L_b =$ Shunt-circuit bridge inductances.

$L_g =$ Inductance of each of the four earth impedances in the bridge.

$L_h =$ Series inductance of C_3.

$(L_h + l) =$ Series inductance of C_1 and C_2.

$L_j =$ Equivalent junction inductance between the 1-bar and 2-bar points.

$L_k =$ Equivalent junction inductance between the 2-bar and 3-bar points.

$L_m =$ Mutual inductance between each of the series elements and its associated earth inductance.

L_{0a}, L_{0b} or $L_0 =$ Total series inductance of C_a or C_b at $\theta = 0$.

$L_{ss} =$ Self-inductance of G_s.

$L_t =$ Total inductance of the terminals.

$l =$ Length, or a small inductance.

R_{0a}, R_{0b} or $R_0 =$ Series resistance of the main variable capacitors C_a or C_b.

r_0 and $r_i =$ Outer and inner radii of a coaxial system.

$R_s =$ Effective series resistance of the bridge resistor.

$Y_a, Y_b =$ Total admittance of the shunt elements of the A- and B-arms.

$\Delta Y_a, \Delta Y_b =$ Change in the admittances Y_a, Y_b.
e.g. $\Delta Y_a = Y_{a2} - Y_{a1}$.

$Y_x =$ Unknown admittance in the plane of the bridge terminals.

Z *General case.* Impedance or characteristic impedance.

        Particular case. Characteristic impedance of the standard coaxial capacitors.

$$Z = Z_\infty(1 + \zeta).$$

$Z_1, Z_2, Z_3, Z_4 =$ Impedance of the bridge series elements C_1, C_2, C_3 and G_s.

$Z_\infty = l/v_\infty C =$ Limiting value of characteristic impedance for zero skin depth.

$Z_c =$ Characteristic impedance of the cone between the bridge terminals and the unknown.

$Z_m = j\omega L_m$.

$Z_g = j\omega L_g$.

$\beta = \omega/v$.

θ_a, θ_b or $\theta =$ Scale readings of the main variable capacitors C_a and C_b.

$v = v_\infty/(1 + \zeta) =$ Velocity of propagation in the standard coaxial capacitors.

$v_\infty =$ Velocity of electromagnetic waves in dry air at $20°$ C and 760 mm Hg ($2\cdot9971 \times 10^{10}$ cm/s), i.e. the limiting velocity in a coaxial system when the skin depth is zero.

$\omega =$ Angular frequency.

$(1 + z) = G_s/G_{ss} =$ Ratio of the r.f. conductance to the d.c. conductance of the bridge resistor.

$(1 + \zeta) = v_\infty/v =$ Reciprocal velocity factor.

$(1 + \sigma) =$ Correction factor for the terminal inductance due to skin effect.

$(1 + \gamma) =$ Incremental inductance correction factor for a conical coaxial line due to skin effect.

(1) INTRODUCTION

Immittance* is the basic parameter of any a.c. measuring system for electrical quantities. In order that any of these quantities may be measured to within a known uncertainty, a knowledge of this basic circuit parameter is required. In the past, the measurement of electrical quantities at radio frequencies to an accuracy within 5 or 10% has been adequate for most applications. This has been accomplished without the need for direct imnittance measurement by making various assumptions; e.g. by employing coaxial cable and assuming the nominal characteristic impedance, and also by employing small carbon resistors and assuming that the d.c. resistance is applicable at high frequencies. These assumptions, whilst they may be justified in certain instances, do nevertheless lead to additional uncertainty in the absolute measurement of other electrical quantities. The uncertainty introduced by these assumptions may be of the order of a few per cent, and they can therefore be ignored when an overall precision of only 10% or so is required. Before any worthwhile improvement can be made in the accuracy of measurement of electrical quantities at high frequencies it is first necessary to devise a means for more accurate immittance measurement. If a precision of 1 or 2% is required in, say, the measurement of voltage and power, the conductance must be known to within a few parts in 10^3 if the concepts of voltage and power are to be compatible so that either one of these quantities can be obtained in terms of the other in accordance with the relationship $P = V^2G$. This requirement rules out the assumptions hitherto adopted, and the direct measurement of immittance becomes a necessity.

The calibration of a source in terms of available power is often employed to overcome the difficulty arising from the lack of reliable means for immittance measurement. A source calibrated in this manner is invariably used with a load incorporating matching facilities. The employment of the principle of conjugate matching inevitably means that the bandwidth of the system is narrow. This gives rise to inconvenience in use and restricts the usefulness and flexibility of the system; furthermore, the concept of voltage is no longer applicable.

With the very wide range of frequencies in current use it is an obvious advantage to have measuring instruments of substantially frequency-independent characteristics, because it avoids the duplication of these instruments for each frequency band. To design such instruments it is essential to have the means for making precise immittance measurements in order to verify the theory. A typical example is the tractorial resistor mount described by Harris.[1] A resistor mount constructed in accordance with the initial theory exhibited an unaccountable shunt capacitance (less than $0\cdot1$ pF) at 200 Mc/s. A fuller theory developed subsequently, which took into account the penetration of the electric field into the resistor body, completely resolved the anomaly.

A further application of the bridge is the calibration of transfer standards for checking the accuracy of less accurate, but less expensive, immittance-measuring instruments. Many other applications will become apparent to the user where high precision is required in the measurement of both the major and minor components.

The problems associated with the measurement of immittance at radio frequencies have received considerable attention in the past. Various methods have been employed, which can be classified broadly under three headings:

     (*a*) Standing-wave-ratio methods.
     (*b*) Resonance methods.
     (*c*) Null methods.

---

* Immittance is used to convey the general idea of impedance or admittance.

Standing-wave-ratio methods are usually restricted to frequencies above one or two hundred megacycles per second, because the length of the slotted line would be excessive at the lower frequencies when, apart from being cumbersome, the manufacturing difficulties would be very great. With this method, optimum performance is usually achieved over the range of impedances associated with coaxial systems, i.e. 45–100 ohms. The best accuracy which can be obtained is of the order of 5% at 100 Mc/s, improving to about 1% for frequencies above 400 Mc/s for a precision slotted line. The range of measurement obtainable is of the order of 10^4–1, i.e. from one hundredth to one hundred times the characteristic impedance of the line.

Resonance methods, in which the unknown immittance is placed either in shunt or in series with a resonant circuit, are usually only suitable when the resistive component of the unknown is either large or small. For resistances in the region of 100 ohms the Q-factor of the resonant circuit is reduced to such a low value that the method becomes impracticable.

Bridges have been designed for operation up to several hundred megacycles per second and to measure the range of immittances associated with coaxial systems, but the inaccuracies are of the order of several per cent. Such instruments invariably use one or more resistors in the bridge circuit. The type of resistor employed is usually selected for low unwanted residual parameters, so that the effective resistance at radio frequency does not depart too widely from the d.c. value. This variation with frequency of the bridge resistance is only one of the many factors which contribute towards the overall inaccuracy of a v.h.f. bridge. It is, however, probably the most difficult problem to be overcome in the design of a precision bridge, since this resistance must be known to within about one part in 10^3.

There appear to be only two possible solutions to the problem: either the *in situ* value of the resistance must be calculable up to the highest frequency used, or the bridge itself must be capable of measuring its own internal resistance in terms of the fundamental units of length and time. The required overall accuracy of the admittance-measuring instrument to be described was a few parts in one thousand. This ruled out the first method because the state of the art in the theoretical treatment of resistors at radio frequency was such that the performance could not be calculated to anything like this accuracy, bearing in mind the difference between a practical design with terminations and a hypothetical resistor such as a solid homogeneous rod or a thin-walled tube with an air core. Ultimately a suitable bridge circuit was found which enabled the second method to be used.

The bridge employs a twin-T circuit. In addition to the well-known merits of this circuit, in which the source, the detector and the unknown have a common earth connection, the circuit selected has a number of other important advantages, namely:

(a) It is amenable to precise mathematical analysis for the effects of unwanted residual parameters provided that a suitable mechanical design is adopted.
(b) It can be used as a dual bridge, and in this way the r.f. conductance of the internal standard resistor of the bridge can be established in terms of length and time measurements.
(c) It provides a range of conductance measurement which is independent of frequency.
(d) It provides means for the compensation of residual effects associated with the series elements and the common earth connection, the effects of which would otherwise influence the fundamental determination of conductance mentioned in (b).

Another special feature of the bridge is a new form of variable air capacitor in which the effective capacitance at 200 Mc/s can be calculated to within 0·2% of the absolute value.

The calibration of the bridge is effected by means of four coaxial standard capacitors whose susceptances are calculated

from length and time measurements. The standards can also be used in the short-circuited condition to provide standards of negative susceptance.

The frequency range of the instrument is from 3 to 300 Mc/s, but the absolute accuracy has, so far, only been assessed at frequencies up to 200 Mc/s. At 200 Mc/s the inaccuracy is 0·2% ± 0·005 pF on capacitance and 0·2% ± 5 micromhos on conductance for admittances of any phase angle. The range of measurement is 50 millimhos and ±50 pF. The discrimination is 1 micromho and 0·001 pF.

It is not possible, within the scope of the paper, to deal with the theory of the instrument in great detail or to describe the techniques employed in the measurement of the residual parameters. However, these are given in Reference 2, which also includes a reference to the manufacturing drawings which are suitable for the reproduction of the instrument elsewhere (see also Reference 7). These drawings embody a major redesign of the mechanism for driving the main variable capacitors in order to ensure better long-term stability of calibration. The original design is beginning to show signs of wear after 10 years' use. The material for all parts carrying r.f. currents has been specified in tellurium-copper. Figs. 4–6 show the original model.

(2) BASIC CIRCUIT

The basic circuit of the bridge is given in Fig. 1. Provision is made for measurements to be undertaken across either arm
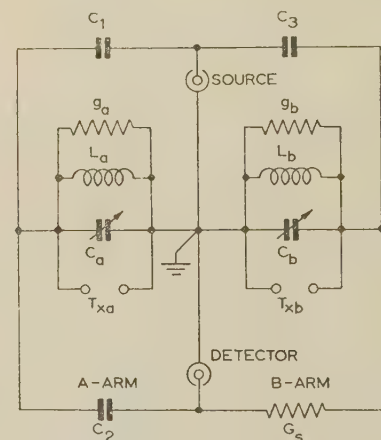
Fig. 1.—Basic circuit of bridge.

of the bridge between the terminals T_{xa} or T_{xb}. This constitutes the dual feature of the instrument. The T_{xb} terminals are for normal measurement purposes, whereas both terminals are needed in order to measure residual effects associated with the resistor G_s. The conductances g_a and g_b represent the combined loss of the shunt elements of each T-network. The inductors L_a and L_b are only needed in order to obtain initial balance; they do not enter into the equations for the unknown admittance.

Neglecting the effects of residuals, the conditions for zero transmission are given by

$$\frac{1}{\omega^2 L_a} = C_a + C_1 + C_2 + \frac{C_1 C_2}{C_3}\left(1 + \frac{g_b}{G_s}\right) \quad . \quad . \quad (1)$$

$$\frac{1}{\omega^2 L_b} = C_b + C_3\left(1 - \frac{g_a G_s}{\omega^2 C_1 C_2}\right) \quad . \quad . \quad . \quad (2)$$

When an unknown admittance $Y_x = G_x + j\omega C_x$ is connected across either arm the components of Y_x are given by

(a) *Unknown across the* B-*arm.*

$$G_{xb} = \frac{G_s C_3}{C_1 C_2} \Delta C_a \quad . \quad . \quad . \quad . \quad (3)$$

$$C_{xb} = \Delta C_b \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where
$$\Delta C_a = C_{a1} - C_{a2}$$

and
$$\Delta C_b = C_{b1} - C_{b2}$$

Subscripts 1 and 2 are used to denote the initial and final values respectively.

(b) *Unknown across the* A-*arm.*

$$G_{xa} = \omega^2 R_s \frac{C_1 C_2}{C_3} \Delta C_b \quad . \quad . \quad . \quad (5)$$

$$C_{xa} = \Delta C_a \quad . \quad . \quad . \quad . \quad (6)$$

where
$$\Delta C_b = C_{b2} - C_{b1}$$

and
$$\Delta C_a = C_{a1} - C_{a2}$$
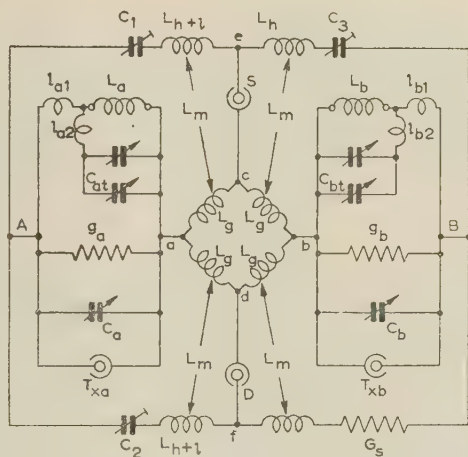
Fig. 2.—Practical circuit of bridge.

C_a: 25 pF (variable)
C_b: 50 pF (variable)
C_1: 5 pF (preset)
C_2: 5 pF (preset)
C_3: 10 pF (preset)

C_{at}: 60 pF and 0·7 pF
C_{bt}: 100 pF and 0·7 pF
L_a and L_b: Plug-in inductors
G_s: 5 millimho plug-in resistor

R_s is the equivalent series resistance of G_s, and will only be equal to the reciprocal of G_s when G_s has zero susceptance. If the equivalent shunt capacitance of G_s is C_{sg},

$$R_s = 1/G_s \left(1 + \frac{\omega^2 C_{sg}^2}{G_s^2}\right). \quad . \quad . \quad . \quad (7)$$

(3) PRACTICAL CIRCUIT

The practical circuit of the instrument is given in Fig. 2. This shows the common earth point split into four separate earth circuits L_g, each mutually coupled to the residual inductance of the four series elements C_1, C_2, C_3 and G_s. The mechanical design is such that L_m is made equal to L_g, and under this condition the four points a, b, c and d are brought to the same potential, thus achieving a common earth point. The residual inductances of C_1, C_2 and C_3 are also proportioned in a certain manner such that $C_1 C_2/C_3$ [eqns. (3) and (5)] is maintained constant up to the highest frequency. In these two ways all the residual effects of the series elements of the bridge are eliminated, except the variation of G_s with frequency. This one remaining variable can then be determined by the dual feature of the bridge.

The two pairs of variable capacitors C_{at} and C_{bt} provide coarse and fine setting adjustments for initial balance when the main capacitors C_a and C_b are set to selected scale values.

The inductors L_a and L_b are of the plug-in type and are used at frequencies up to 180 Mc/s. Above this frequency it was not possible to construct inductors of a sufficiently low value to obtain initial balance. In order to overcome this difficulty, a series circuit was adopted in order to obtain the large negative susceptance. This is accomplished by means of the four inductors l_{a1}, l_{a2}, l_{b1} and l_{b2}, which are built into the bridge. These, in conjunction with C_{at} and C_{bt}, provide a means for obtaining a variable negative susceptance between $A - a$ and $B - b$. The two series circuits are, of course, operated above resonance in order to obtain this condition. They work over the range 180–300 Mc/s.

(4) EQUIVALENT CIRCUIT

The equivalent circuit of the bridge, showing all the significant residual parameters, is given in Fig. 3. The inductance and distributed capacitance of the terminals are shown as L_t and C_t, and the terminal fringe capacitance as C_f. The residual parameters of the
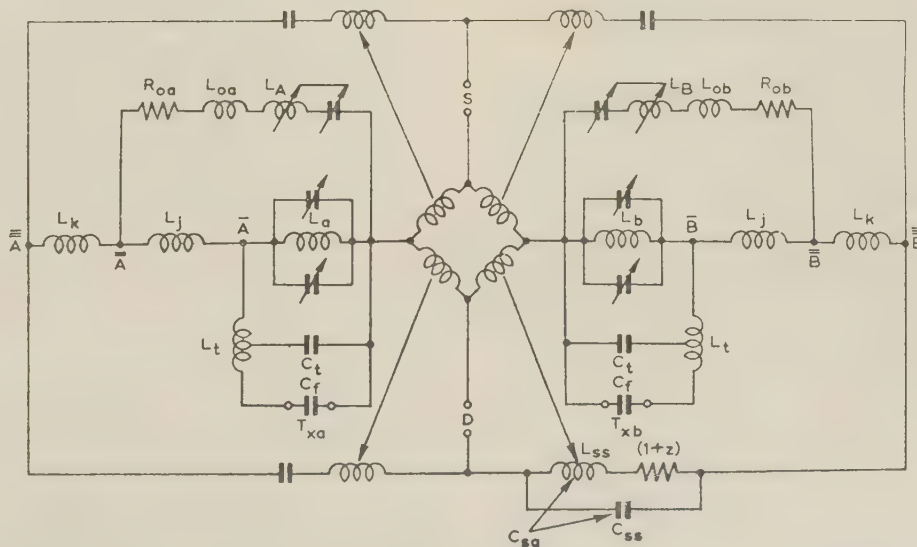
Fig. 3.—Equivalent circuit of bridge.

main variable capacitors C_a and C_b are shown as R_{0a}, L_{0a}, L_A and R_{0b}, L_{0b}, L_B, respectively. The residual inductance of each of these capacitors comprises a fixed inductance and a variable portion which is ganged to the capacitor. The laws of the variable inductor and the capacitor are both linear but vary in opposite senses. This form of capacitor is amenable to precise circuit analysis and enables the effective capacitance to be calculated very accurately up to several hundred megacycles per second. The equivalent circuit of the junction between the shunt and series elements is shown as L_j and L_k for each T-network. C_{ss} is the capacitance across G_s, and L_{ss} is the self-inductance which is mutually coupled to its associated earth inductance. The combined effect of C_{ss}, L_{ss} and the mutual inductance can be represented by an effective shunt capacitance C_{sg}. The factor $(1 + z)$ represents the ratio of the r.f. conductance to the d.c. conductance.

The bar markings over the A and B symbols indicate the 1-bar, 2-bar and 3-bar points. Between these points changes in the admittances of the shunt circuit elements have to be transformed in order to express all admittance changes in terms of 3-bar values, so that they can be related to the bridge equations containing the series elements. These changes are caused by the connection of the unknown admittance to the terminals and the subsequent adjustment of C_a and C_b for final balance. The adjustment of these two capacitors causes complex admittance changes at high frequencies owing to their resistances.

The two inductors L_a and L_b, required for initial balance purposes, are each shown connected in parallel with their associated trimmers, the combinations being connected by single conductors to the 1-bar points. This condition results in a very great simplification in the equivalent circuit and is achieved by placing each inductor and its associated trimmers in a separate cavity wherein they are connected together, the combination being connected by a coaxial link-line to the other bridge elements. The inductances of these link lines are shown as l_{a1} and l_{b1} in Fig. 2.

(5) GENERAL DESCRIPTION

A front view of the bridge in its final form is shown in Fig. 4. The two large micrometer heads at the back drive the main
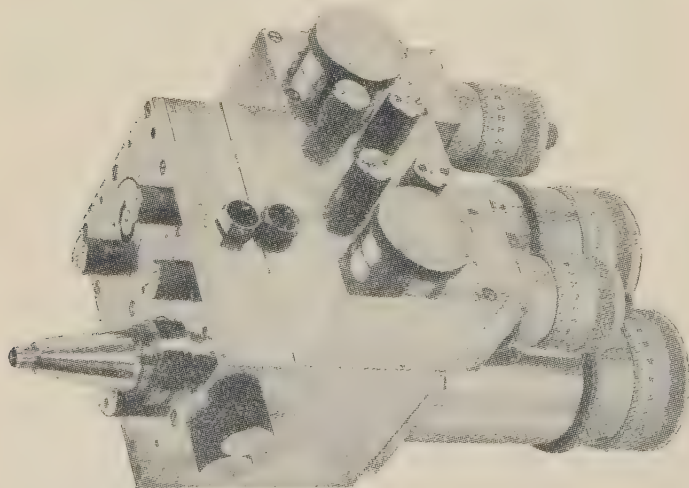
Fig. 4.—View of bridge.

capacitors C_a and C_b. The smaller micrometer heads above are for the fine trimmers, whereas the coarse trimmers are controlled by means of the two large knobs on the top of the instrument. The smaller knobs on this face are the plug-in inductors, either one inductor or two inductors in parallel being used for each

T-network. In this way the total number of inductors required for the complete frequency range is considerably reduced. The two knobs on the right-hand side of the instrument are removable caps which give access to C_3 and G_s. Two similar caps on the opposite face provide access to C_1 and C_2. The input plug is clearly shown and is directly opposite the output plug on the bottom face of the instrument. The front face embodies the coaxial dual terminal system. The B-arm terminals are shown with a cone fitting attached and the A-arm terminals open-circuited. The four caps on this face provide access to the main junctions, so that external capacitors can be used at low frequencies to increase the range of susceptance measurement.

It will be seen that the instrument is constructed from two blocks, the primary block with the secondary block on top. The former contains all the circuit elements which enter into the bridge equations for the unknown. These circuit elements are called 'dependent' elements. The remaining circuit elements are located in the secondary block and are called 'independent' elements. The independent circuit elements for each T-network are connected in parallel inside the secondary block for the reason already given.
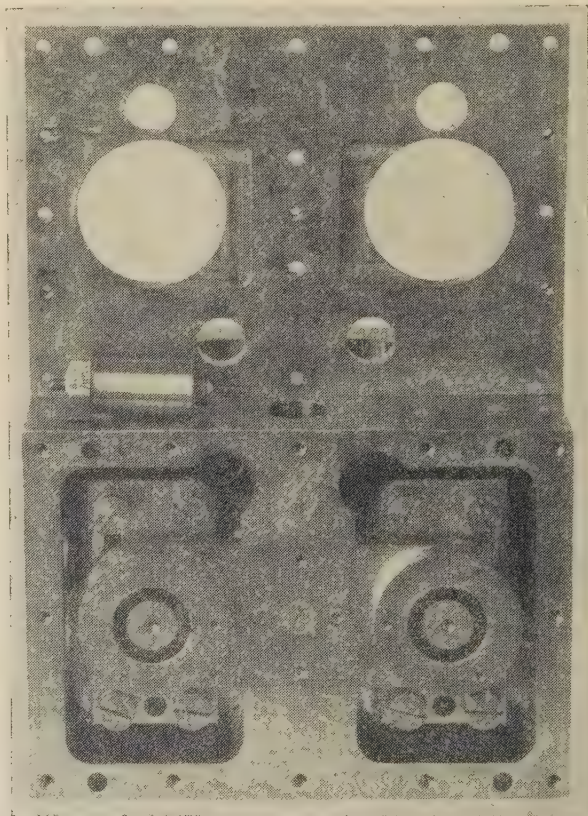
Fig. 5.—View of bridge with front plate removed.

Fig. 5 is a view with the front plate lifted off, showing the internal construction of the dual-terminal system and the manner in which the common impedance between the low-potential electrode of the terminals and the two cavities is reduced. The cavities, apart from forming the earth-return circuits of C_a and C_b, also house the main junction blocks (A and B in Fig. 2) which are supported on fused-quartz washers. These blocks carry the electrodes of all the dependent circuit elements, as well as the high-potential electrodes of the terminals. The web between the cavities houses the four series elements C_1, C_2, C_3 and G_s.
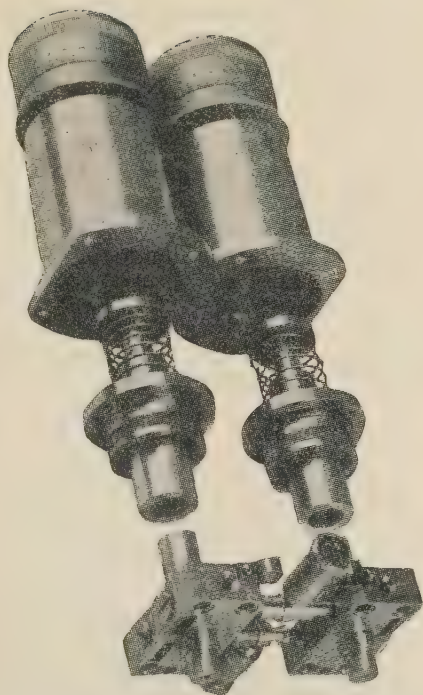
Fig. 6.—View of dependent circuit elements and exploded view of main variable capacitors.

Fig. 6 shows all the dependent circuit elements removed from the primary block and reassembled in their relative positions, except the main variable capacitors, of which an exploded view is given. The coaxial construction adopted for the series capacitors, which are mounted inside the web, can be clearly seen between the two junction blocks, together with the plug-in resistor G_s in the lower right-hand position. The two vertical cylinders between the junction blocks are the source and detector coaxial connections, the inner conductors of which plug into the centres of the two horizontal cylindrical junction-pieces of the four series elements. A better appreciation of the construction of this part of the instrument can be obtained from Fig. 7. Behind the junction blocks are the cylindrical series electrodes of the variable capacitors C_a and C_b. These cylinders are mounted on fused-quartz washers and are electrically floating. The quartz washers are mounted on spring-loaded slides which are driven by the micrometer heads. The stubs protruding behind the junction blocks are the high-potential electrodes of these capacitors. Each of these electrodes is located by means of a coaxial fused-quartz washer which spigots into a cylindrical hole in the primary block at the bottom of a main cavity and which forms the low-potential electrode. The cylinder is driven into the annular gap between these electrodes. This form of construction eliminates rubbing contacts and produces a linear capacitance law as well as a linear inductance change with dial setting. It will be realized that each capacitor is formed by two air-gaps in series. The widths of these gaps are $0 \cdot 004$ in for C_b and $0 \cdot 0075$ in for C_a, and the machining calls for a very high order of mechanical precision. The length of traverse of both capacitors is approximately $0 \cdot 8$ in.

Fig. 7 is a section through the main cavities in the plane of the series elements. The Figure illustrates how the earth network of four inductances L_g is formed and the manner in which the four series elements are mutually coupled to their respective earth circuits, whilst, at the same time, the mutual inductances are made substantially equal to the earth inductances. It will be seen that junction 7, between C_1, C_3 and the input, is mounted
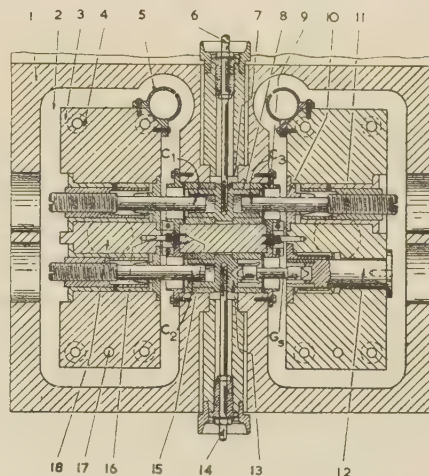
Fig. 7.—Section through the main cavities in the plane of the series elements.

1. Primary block.
2. A-arm cavity.
3. A-arm main junction block.
4. Quartz washers supporting 3.
5. Socket for external susceptance capacitor.
6. Input plug.
7. Junction between C_1, C_3 and input.
8. Quartz insulators supporting 7.
9. Clamping rings, incorporating screens, for 7 and 8.
10. Phosphor-bronze contact for adjustable electrode of C_3.
11. Micrometer head for C_3.
12. Plug-in resistor chuck.
13. Junction between C_2, G_s and output.
14. Output plug.
15. Connection between link-line outer and A-arm cavity.
16. Connection between link-line inner and A-arm junction block.
17. Alternative socket to 5.
18. High-potential electrode of C_a.

in a cylindrical hole in the web between the two cavities. The surface of this hole forms two of the earth inductances, with the centre corresponding to the point c (Fig. 2). The point a corresponds to the centre of the right-hand face of the left-hand cavity (15 in Fig. 7). The point b in the earth network lies in the centre of the left-hand face of the right-hand cavity. Similarly the point d is the centre of the other hole in the web housing the junction of C_2 and G_s. The centres of the two junctions 7 and 13 are the points e and f, respectively. Because the series elements are mounted inside these holes, mutual inductance will exist between each series element and its associated earth path. The mutual inductance between two coaxial conductors is always equal to the inductance of the outer. Therefore equality of mutual inductance L_m and earth inductance L_g is automatically achieved merely by placing the series elements inside the earth circuits.

(6) TERMINALS

In the design of the terminals of a precision immittance-measuring instrument a number of important factors must be taken into account if the accuracy of the instrument is not to be impaired by an uncertainty in the discontinuity at the junction between the terminals and the connector attached thereto. The precision of the bridge under discussion is such that it is capable of measuring admittance in a uniform coaxial system of substantially zero voltage-standing-wave ratio with an inaccuracy corresponding to an uncertainty of $0 \cdot 0015$ in in the position of the plane of measurement.

The terminals must also be suitable for the attachment of standards, and the method of connection must be such that the calculated values are not changed by a significant amount, i.e. a few parts in 10^4.

There is no doubt that the terminals should be coaxial in form. In order to decide on a suitable characteristic impedance it is necessary to consider the problem of immittance measurement generally. In order to cover the complete range of immittance measurement by bridge methods it is necessary to employ both impedance and admittance bridges. For any given frequency range the two bridges are complementary to one another. In order that the errors due to terminal immittance shall be small the impedance bridge requires a low terminal admittance and the admittance bridge a low terminal impedance. If, therefore, it is desired to adopt a standardized design of terminal for both types of bridge, so that a common set of immittance standards can be used, a compromise must be made in the value of the terminal immittance. It can be shown that this optimum value is that which gives rise to equal errors in both types of bridge. It will fall in the immittance region where the ranges of the two types of bridge overlap. A consideration of these factors leads to the result that this optimum impedance is in the region of 25–50 ohms. The terminal diameter ratio chosen was $1 \cdot 500$, corresponding to an impedance of $24 \cdot 3$ ohms. The dimensions of the outer and inner conductors are $0 \cdot 750\,0$ in and $0 \cdot 500\,0$ in, respectively. The outer termination is a flat earth-plate, and the inner termination is a circular stub flush with the earth plane. The dielectric is air.

In addition to the above requirements the terminals must be suitable for the attachment of connectors which form part of other measuring instruments in which immittance is a major parameter, e.g. wattmeters, attenuators, etc. The bridge terminal, briefly described here, forms part of a universal coaxial-connector system for precision radio-frequency measuring instruments of all types.[3]

In order that the plane of measurement shall be defined by the physical plane of the bridge termination, it is necessary to take into account the fringe field at the open terminals of the admittance bridge. This fringe field has been determined by two methods with the following result:

$$C_f = 0 \cdot 343 \text{ pF} \pm 1\% \quad . \quad . \quad . \quad . \quad (8)$$

The total terminal inductance and capacitance were measured in terms of the susceptance standards at a frequency of 200 Mc/s with the following results:

$$L_t = 2\,750 \text{ pH} \pm 0 \cdot 5\% \quad . \quad . \quad . \quad (9)$$

$$C_t = 4 \cdot 6 \text{ pF} \pm 2\% \quad . \quad . \quad . \quad . \quad (10)$$

The terminals form the link between the unknown admittance and the 1-bar points (Fig. 3). The bridge determines the unknown admittance at the 1-bar points. This admittance has then to be transformed through the terminal impedance. The following formulae have been developed for this purpose because other methods, such as a circle diagram, are not sufficiently accurate:

$$G_x = \frac{\bar{G}_x(1 + \omega^2 L_t C_t)}{[1 + \omega^2 L_t(\bar{C}_x + C_t)]^2 + \omega^2 L_t^2 \bar{G}_x^2} \quad . \quad (11)$$

$$C_x = \frac{\bar{C}_x[1 + \omega^2 L_t(\bar{C}_x + C_t)] + (1 + \sigma)L_t \bar{G}_x^2}{[1 + \omega^2 L_t(\bar{C}_x + C_t)]^2 + \omega^2 L_t^2 \bar{G}_x^2} + \omega^2 L_t C_f(C_f + C_t)$$
$$. \quad . \quad . \quad (12)$$

where $(1 + \sigma)$ is a factor depending upon the skin depth in the terminals. This factor is only significant at frequencies below 100 Mc/s. The above formulae give the components of the unknown admittance in the plane of the bridge terminals for an undistorted radial field in that plane.

(7) THE SUSCEPTANCE STANDARDS

When the practical design of the bridge was commenced in 1947 no standards of immittance were available. The state of the art in the design of conductance standards was such that the required order of accuracy of $0 \cdot 1\%$ at 200 Mc/s could not be realized. The design of coaxial susceptance standards was possible, however, provided that careful attention was given to the problem of terminations. This meant that the bridge itself must be capable of determining the true conductance of its own internal conductance standard in terms of external susceptance standards. This was the main requirement which led to the dual feature of the bridge.

A sectional view of one of the standards connected to the bridge terminals is given in Fig. 8. This comprises two coaxial
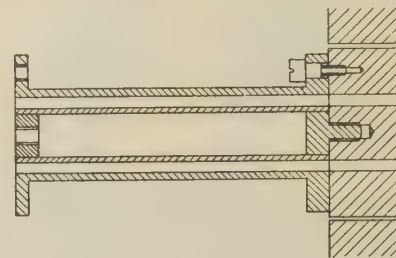
Fig. 8.—Section through a standard capacitor and the bridge terminals.

copper tubes which are self-supporting on the bridge terminals and of the same effective diameters. The outer electrode is secured by three dowel screws, and the inner electrode is screwed into the terminal inner conductor. The remote end of the capacitor is a replica of the bridge terminals and thus produces the same open-circuit conditions as exist at the bridge terminals. The change in capacitance produced at low frequencies is therefore equal to that caused by an undistorted radial field. There are four capacitors having values of $2 \cdot 5$, 5, 10 and 20 pF. These can be used singly or combined with one another in twos and threes to produce standards of capacitance in steps of $2 \cdot 5$ pF up to 35 pF.

The capacitance was calculated from the following formula:

$$C = \frac{0 \cdot 241\,74\,l}{\log_{10}\dfrac{r_0}{r_i}} \text{ pF, for dry air at 20° C and 760 mm Hg}$$
$$. \quad . \quad . \quad (13)$$

The constant is based on the velocity of electromagnetic waves in vacuo as $2 \cdot 997\,93 \times 10^{10}$ cm/s and the dielectric constant of air as $1 \cdot 000\,53$. The uncertainty in the calculated values based on length measurements did not exceed $0 \cdot 1\%$ $\pm 0 \cdot 005$ pF.

At high frequencies the effective capacitance C' of a substantially lossless coaxial capacitor, with no fringe field at the open end, can be expressed in terms of the static capacitance C by the following relationship:

$$C' = C\frac{\tan \beta l}{\beta l} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

where

$$\beta = \omega/v$$

$$v = v_\infty/(1 + \zeta)$$

The factor $(1 + \zeta)$ arises from the change in inductance due to finite current penetration at relatively low frequencies. This affects both the velocity of propagation and the characteristic impedance but not the capacitance C. The characteristic

impedance at high frequencies, when the skin depth is substantially zero, is given by

$$Z_\infty = 138 \cdot 02 \log_{10} \frac{r_0}{r_i} \text{ ohms} \quad . \quad . \quad . \quad (15)$$

or

$$Z_\infty = l/v_\infty C \quad . \quad . \quad . \quad . \quad . \quad (16)$$

At low frequencies, where the skin depth is significant,

$$Z = Z_\infty/(1 + \zeta) \quad . \quad . \quad . \quad . \quad (17)$$

or

$$Z = l/vC \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The factor $(1 + \zeta)$ is called the 'reciprocal velocity factor', and the calculated values are given in Fig. 9 for a coaxial system having the same dimensions as the standard capacitors.
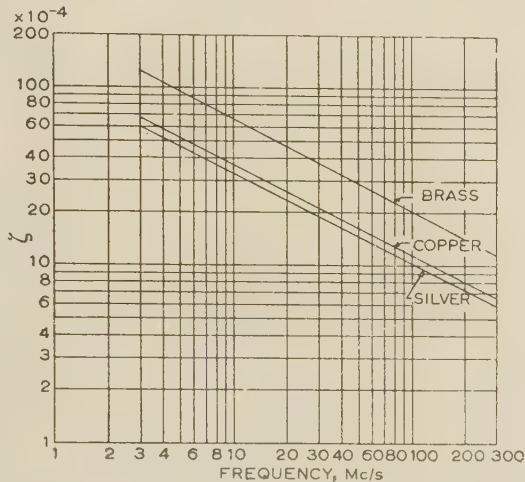
Fig. 9.—Reciprocal velocity factor $(1 + \zeta)$ for a coaxial line having diameters of $0 \cdot 750$ and $0 \cdot 500$ in.

At high frequencies the fringe field at the open end of the capacitor will produce a greater effect than the same fringe field that exists at the bridge terminals before attachment of the capacitor. If the fringe field is called C_f it can be shown that the input capacitance of the capacitor terminated by C_f will be

$$C'' = \frac{C' + C_f}{1 - \dfrac{C_f}{C'} \tan^2 \beta l} \quad . \quad . \quad . \quad (19)$$

The change in capacitance at the bridge terminals due to the connection of the capacitor will therefore be

$$\Delta C'' = C'' - C_f \quad . \quad . \quad . \quad . \quad (20)$$

In order to provide a range of negative susceptance standards a short-circuit in the form of a metal diaphragm is applied at the open end of the capacitor. The inductance of such a short-circuit has been measured by means of the bridge and found to be less than 3 pH, which is negligibly small. The short-circuited standards can be built up in various values in the same manner as the positive susceptance standards. The effective input capacitance of a short-circuited lossless coaxial line is given by

$$C'_n = - 1/\omega^2 Z^2 C' \quad . \quad . \quad . \quad . \quad (21)$$

The change in capacitance at the bridge terminals due to connection of the standard will be

$$\Delta C'_n = C'_n - C_f \quad . \quad . \quad . \quad . \quad (22)$$

The fringe capacitance at the open end of the standards, and at the bridge terminals, has associated with it a small con-

ductance caused by radiation. This varies as the fourth power of the frequency, but it is negligibly small and amounts to only $0 \cdot 017$ micromho at 300 Mc/s.

(8) VARIABLE CAPACITORS C_a AND C_b

For a variable capacitor to be suitable for use in a v.h.f. bridge, either its residual inductance must be negligibly small or this inductance must follow a well-defined law and in such a manner that the equivalent circuit is amenable to precise mathematical treatment. Also, it must be possible to devise experiments so that this residual inductance can be measured to the required order of accuracy.

For a capacitor of 50 pF at a frequency of 200 Mc/s the inductance which would cause an error of $0 \cdot 1\%$ is $12 \cdot 5$ pH. The realization of this condition in practice was not possible, and so a capacitor of new design was developed. In form it is coaxial, with a double air-gap between a moving insulated cylindrical electrode and two fixed coaxial electrodes. This results in the capacitance and inductance changes both bearing a linear relationship to the dial setting, but varying in opposite senses.

A diagrammatic representation of the capacitor is given in Fig. 10, where P is the unsupported end of the moving electrode.
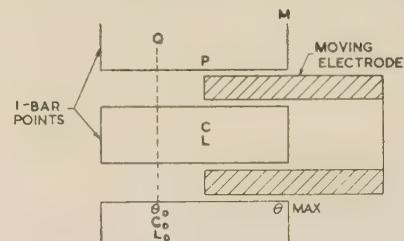
Fig. 10.—Diagrammatic representation of C_a and C_b.

The traverse of this electrode is from Q at maximum capacitance, when $\theta = 0$, to M at minimum capacitance. As θ increases the capacitance decreases linearly at, say, a rate C and the inductance increases linearly at, say, a rate L. If the capacitance and inductance are C_0 and L_0 at θ_0, the values at any setting θ will be $(C_0 - C\theta)$ and $(L_0 + L\theta)$, respectively. At high frequencies the effective capacitance at the 1-bar points will be

$$\bar{C} = \frac{C_0 - C\theta}{1 - \omega^2(L_0 + L\theta)(C_0 - C\theta)} \quad . \quad . \quad (23)$$

Fig. 11 shows how the capacitance varies with dial setting at high frequencies. It will be noted that a point of inflection occurs
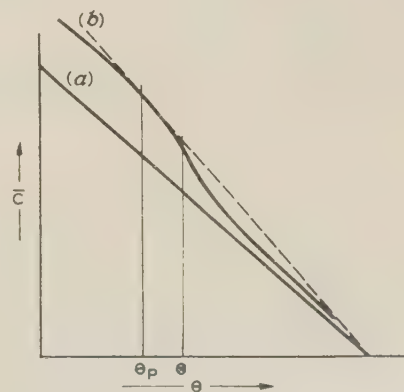
Fig. 11.—Effective capacitance of C_a and C_b.
(a) Low-frequency calibration.
(b) High-frequency calibration.

at Θ. Here $d\bar{C}/d\theta$ is a maximum, so that if a small fixed capacitance $\Delta\bar{C}$ is measured across the 1-bar points throughout the range of the capacitor, the corresponding values of $\Delta\theta$ will reach a minimum at this point. By suitable design this point can be made to occur at approximately the mid-scale value in order to ensure the best possible fit between the experimental results and the theoretical curve when measuring the residual parameters. It can be shown by double differentiation of eqn. (23) that both L and L_0 can be expressed in terms of $\Delta\bar{C}$ and the minimum value of $\Delta\theta$, namely $\Delta\Theta$:

$$L = \frac{C}{\omega^2(C_0 - C\Theta)^2}\left(1 - \sqrt[3]{\frac{-C\Delta\Theta}{\Delta\bar{C}}}\right). \quad . \quad . \quad (24)$$

$$\frac{L_0}{L} = \frac{(C_0 - C\Theta)}{C}\left(1 + \sqrt[3]{\frac{-C\Delta\Theta}{\Delta\bar{C}}}\right) - \Theta \quad . \quad . \quad (25)$$

These expressions are obtained without approximations. When L and L_0 have been determined experimentally and substituted back in eqn. (23) a formula is obtained which enables the effective capacitance at 200 Mc/s to be calculated to within 0·1% of the values given by the susceptance standards.

The measurement of the residual parameters L and L_0 was carried out at a frequency of 200 Mc/s by means of a coaxial capacitor of approximately 0·5 pF, comprising an annular ring of 0·750 0 in internal diameter and a disc of 0·500 0 in diameter, each 0·15 in thick.

The experimental values obtained for the residual parameters were as follows:

  (a) Capacitor C_a
$$L_A = 3\,642\,\text{pH/in} \pm 0·5\%$$
$$L_{0a} = 2016\,\text{pH} \pm 0·5\%$$

  (b) Capacitor C_b
$$L_B = 3\,353\,\text{pH/in} \pm 0·5\%$$
$$L_{0b} = 1\,874\,\text{pH} \pm 0·5\%.$$

The accuracy of the formulae for $\bar{C}_a$ and $\bar{C}_b$ was subsequently verified at various frequencies by means of the negative and positive susceptance standards. Each of these standards was measured over different parts of the scales of C_a and C_b. In no case did the difference exceed 0·1% ± 0·005 pF.

The initial calibration of C_a and C_b was effected by means of a precision Schering bridge at a frequency of 1 kc/s. This calibration was then verified by means of the positive susceptance standards with the dual admittance bridge operating at 4·5 Mc/s. At this frequency the effect of the residual inductance of the capacitors and the standards is less than one part in 10^4. The 1 kc/s calibration of C_a and C_b agreed to within 0·1% ± 0·005 pF of the calculated values of the standards.

This form of variable capacitor possesses a number of other interesting characteristics. If eqn. (25) is rearranged we get

$$\sqrt[3]{\frac{-C\Delta\Theta}{\Delta\bar{C}}} = \frac{\left(\dfrac{L_0}{L} + \Theta\right)}{\left(\dfrac{C_0}{C} - \Theta\right)} - 1 \quad . \quad . \quad (26)$$

$(L_0/L + \Theta)$ is the distance between the 1-bar points and Θ, and $(C_0/C - \Theta)$ is the distance between Θ and the remote end of the capacitor, where $C_0 - C\theta$ is zero. These relationships are shown in Fig. 12. At very low frequencies the cube-root term will be unity, from which it is apparent that the point of inflection first appears when $(L_0/L + \Theta) = 2(C_0/C - \Theta)$. It will thus be seen that, in the design of a capacitor of this type where the point of inflection is required to fall near the mid-
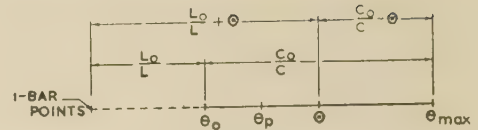
Fig. 12.—Relationship between the parameters of C_a and C_b at high frequencies.

scale value, the length of the variable portion C_0/C should be made twice the length of the portion L_0/L connecting the capacitor to the plane of measurement. As the frequency is increased this point of inflection moves slowly to the left. For example, in the case of C_a, where $\theta_{max} = 0·85$ in, this point first appears at $\theta = 0·44$ in. At frequencies of 200 and 300 Mc/s it occurs at 0·43 and 0·417 in, respectively.

It can also be seen from eqn. (23) that the maximum percentage correction at any frequency always occurs when $(L_0 + L\theta)(C_0 - C\theta)$ is a maximum. The value of θ for this condition is given by

$$\theta_p = \frac{1}{2}\left(\frac{C_0}{C} - \frac{L_0}{L}\right) \quad . \quad . \quad . \quad . \quad (27)$$

If $C_0/C = 2L_0/L$ as previously stated, θ_p occurs at a value of $C_0/4C$ (Figs. 11 and 12). The lowest frequency at which the capacitor exhibits series resonance will, of course, occur at $\theta = \theta_p$. The point of inflection, therefore, always falls somewhere between one-half and one-quarter of the total traverse of the moving electrode. As already stated, this point moves slowly at first with increasing frequency, but as series resonance is approached, it moves rapidly until it coincides with θ_p at resonance. The resonant frequencies of C_a and C_b at θ_p are approximately 632 and 484 Mc/s, respectively.

The resistances of C_a and C_b are significant at the higher frequencies. The effect is to cause a conductance change which must be allowed for when applying corrections for the residual parameters generally. It can be shown that the conductance change is

$$\Delta G = \omega^2 R_0(\bar{C}_2^2 - \bar{C}_1^2) \quad . \quad . \quad . \quad (28)$$

where R_0 is the series resistance, and the effective capacitance change is $\bar{C}_2 - \bar{C}_1$.

The resistance can be determined by measuring a relatively lossless capacitance over different portions of the range of the variable capacitor. The experimental value obtained for both C_a and C_b at 200 Mc/s was

$$R_0 = 0·03\,\text{ohm} \pm 10\% \quad . \quad . \quad . \quad (29)$$

The material used for the electrodes of C_a and C_b is brass. The manufacturing drawings now specify tellurium-copper, which will result in the resistance being reduced to one-half the value in eqn. (29).

(9) SERIES CAPACITORS C_1, C_2 AND C_3

It is necessary to eliminate any error associated with the series elements of the bridge in order that the true conductance of G_s can be established. It was not possible to design the three series capacitors so that their individual values did not change with frequency. This does not matter, provided that C_3/C_1C_2 can be maintained constant, because it is this quantity which appears in both equations for the A- and B-arm methods of connection [eqns. (3) and (5)].

This condition is accomplished by making $C_1 = C_2 = \frac{1}{2}C_3$ and the inductance of C_1 and C_2 equal but slightly greater than that of C_3. If the effective value of C_3/C_1C_2 is K_3/K_1K_2, then, using the nomenclature of Fig. 2,

$$\frac{K_3}{K_1 K_2} = \frac{C_3}{(1 - 2\omega^2 L_h C)} \frac{[1 - \omega^2 (L_h + l)C]^2}{C_1 C_2}$$

$$= \frac{C_3}{C_1 C_2} \left[1 + \frac{\omega^4 (L_h + l)^2 C^2 - 2\omega^2 l C}{1 - 2\omega^2 L_h C} \right] \quad . \quad (30)$$

where $C = C_1 = C_2 = \frac{1}{2} C_3$.

It will be seen that, by proportioning the values in this manner, all first-order ω^2 terms containing L_h cancel. The small residual ω^4 term is compensated over a limited frequency range at the extreme frequencies by the small ω^2 term containing l, which is the difference in inductance between C_3 and the other two capacitors.

The error term in eqn. (30) is shown plotted in curve (a) of Fig. 13. Curve (b) is for the same term when $l = 0$, and it
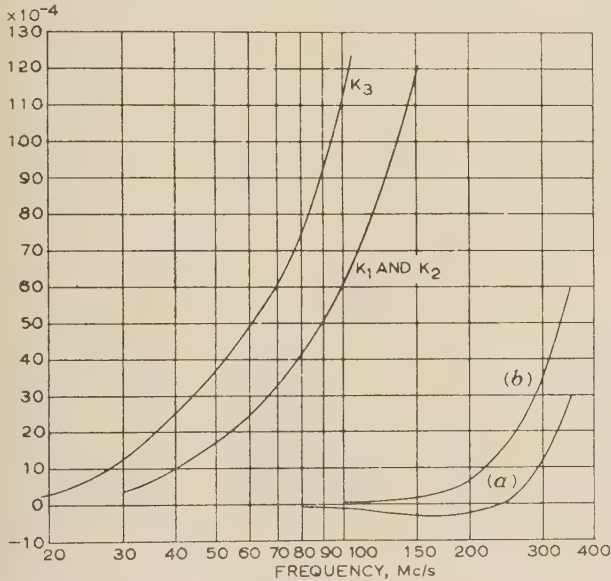
Fig. 13.—Compensation of $C_3/C_1 C_2$.

shows the improvement obtained at the extreme frequencies by making the inductance of C_1 and C_2 greater than that of C_3 by this amount. The errors in the individual values of K_1, K_2 and K_3 are also shown for comparison, from which it will be seen that they are approximately one hundred times greater than the errors in $K_3/K_1 K_2$. The value of L_h is $3\,000\,\text{pH}$ and l is $50\,\text{pH}$. The mechanism of compensation is dependent upon L_h being the same for all three capacitors, and this is realized in the mechanical design by symmetry alone, the actual value being of secondary importance. The small difference in inductance between that of C_3 and the other two capacitors is obtained by suitably dimensioning the coaxial electrodes of this capacitor.

The individual values of the three capacitors were first adjusted to within $0 \cdot 1\%$ of their nominal values by means of a three-terminal Schering bridge at a frequency of $1\,\text{kc/s}$. The overall value of $C_1 C_2/C_3$ was then derived in terms of C_a and a d.c. conductance ratio by means of the dual admittance bridge at a frequency of $4 \cdot 5\,\text{Mc/s}$, using the relationship given by eqn. (3):

$$\frac{C_1 C_2}{C_3} = \frac{G_s}{G_{xb}} \Delta C_a \quad . \quad . \quad . \quad . \quad (31)$$

The resistor used for G_{xb} in this experiment was similar to G_s and consisted of a small carbon-film type, unspiralled, mounted in a coaxial jig. At this frequency the r.f. conductance did not depart from the d.c. value by more than 3 parts in 10^5. The

uncertainty in the determination of $C_1 C_2/C_3$ by this method did not exceed 3 parts in 10^4.

(10) EARTH JUNCTION IMPEDANCES

The presence of any common or mutual impedance between the four earth connections (Fig. 1) will introduce impedance in series with each T-network and give rise to errors in admittance measurement. The elimination of this form of error is also necessary in order that G_s may be determined by means of the dual bridge. A common inductance of the order of $2\,000\,\text{pH}$ is sufficient to cause an error of several per cent at $200\,\text{Mc/s}$. It was not possible to realize a practical design in which these residual effects could be kept to a sufficiently low value. This difficulty was overcome by a new type of earth circuit shown in Fig. 2. The general circuit of the bridge including this earth network is given in Fig. 14, in which Y_a and Y_b represent the
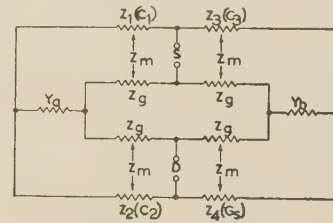
Fig. 14.—General circuit.

combined admittance of the shunt elements of the A and B-arms, respectively.

In order to find a general equation for determining an unknown admittance in terms of the bridge parameters, it is necessary to find the relationship between a change in Y_a in terms of a change in Y_b between two sets of conditions corresponding to zero transmission, with and without the unknown connected across either Y_a or Y_b. This is called the 'balance difference equation', and a rigorous solution to the problem gives the following result:

$$\Delta Y_a \left[Z_1 Z_2 - \frac{Z_m^2}{Z_g}(Z_1 + Z_2) + Z_m^2 \left(\frac{2Z_m}{Z_g} - 1 \right) \right]$$

$$+ \Delta Y_b \left[Z_3 Z_4 - \frac{Z_m^2}{Z_g}(Z_3 + Z_4) + Z_m^2 \left(\frac{2Z_m}{Z_g} - 1 \right) \right]$$

$$+ \frac{(Z_g - Z_m)^2}{4Z_g} [(\Delta Y_a + \Delta Y_b)(Z_1 + Z_2 + Z_3 + Z_4)$$

$$+ (\Delta Y_a Y_{b1} + \Delta Y_b Y_{a1} + \Delta Y_a \Delta Y_b)(Z_1 - Z_3)(Z_2 - Z_4)]$$

$$= 0 \quad . \quad (32)$$

Where Y_{a1} and Y_{b1} are the initial values of Y_a and Y_b, respectively, which include, of course, the admittances of the inductors L_a and L_b and the trimmers C_{at} and C_{bt}. The corresponding balance difference equation for the simple case with a common earth point is

$$\Delta Y_a Z_1 Z_2 + \Delta Y_b Z_3 Z_4 = 0 \quad . \quad . \quad . \quad . \quad (33)$$

It will be seen from eqn. (32) that when common earth impedance is present, i.e. when $Z_m = 0$, a large number of error terms arise from the third main bracket, the common factor of which is then $Z_g/4$. The first two main brackets reduce to the terms in eqn. (33). When mutual impedance Z_m is introduced and made equal to the earth impedance Z_g, the common factor of the third main bracket is zero and the rest of the equation reduces to

$$\Delta \overline{\overline{Y}}_a (Z_1 - Z_g)(Z_2 - Z_g) + \Delta \overline{\overline{Y}}_b (Z_3 - Z_g)(Z_4 - Z_g) = 0$$

$$. \quad . \quad . \quad (34)$$

This is the same as eqn. (33), except that the residual impedance
of each of the four series elements is reduced by an amount
equal to the earth impedance. This is an even better condition
than the simple case with a common earth point. Eqn. (34) is
the general balance difference equation of the bridge for the
3-bar values of ΔY_a and ΔY_b.

In practice, it is unlikely that the mutual inductance L_m will
be exactly equal to the earth inductance L_g. However, con-
siderable departure from this ideal condition can be tolerated
without affecting the degree of compensation by a significant
amount. The factor in question can be written as

$$\frac{Z_g}{4}\left(1 - \frac{Z_m}{Z_g}\right)^2$$

from which it can be seen that, if $Z_m/Z_g = 0.9$, the errors will
be one hundred times less than when $Z_m = 0$. There is experi-
mental evidence that Z_m is not appreciably less than $0.9Z_g$. It
should also be noted that no error terms arise from the first
two main brackets of eqn. (32) for any value of Z_m/Z_g. The
amount by which the impedance of the four series elements is
reduced is modified, however, but they are all affected equally
because of the symmetry of the mechanical design, and so it has
no significant effect on the compensation of C_1C_2/C_3.

It has already been explained how the errors associated with
the three series capacitors have been reduced to a comparable
degree by compensation. In order to obtain a similar per-
formance without compensation, the inductance of the series
capacitors and the earth paths would have to be reduced one
hundredfold—in other words the linear dimensions of these
circuits would have to be reduced by this factor. In the bridge
described these circuits occupy a volume of approximately 1 in³,
so that a bridge of comparable performance, without means for
compensation, would have to be contained within a volume of
about 10^{-6} in³.

(11) MAIN-JUNCTION BLOCK IMPEDANCE

The main junctions of the A- and B-arm circuits consist of
rectangular brass blocks housed in separate cavities. These
blocks form the junction between the series and shunt elements
and the unknown terminals of each T-network. The equivalent
circuit of these junctions is shown as L_j and L_k in Fig. 3.

The equivalent circuit of the junction between the series and
shunt elements of a certain twin-T bridge is given in Reference 4.
In this representation a single inductance was used, i.e. Fig. 3,
in which L_j is zero. It was found that when the experimental
results, relating to the measurement of residual effects, were
fitted to such a circuit the bridge gave anomalous results for
certain types of measurement. It was eventually established that
the equivalent circuit of the junction must be considered as a
series inductance with the individual shunt elements tapped
along it in any combination. There are three shunt elements in
the bridge—the terminals for connecting the unknown admittance,
the main variable capacitor and the combined effect of the
inductor and the two trimmers.

Fig. 15(a) is a representation of this circuit, in which the three
shunt elements are shown as X, Y and Z. The end portion, L_z,
of the equivalent junction inductance can be considered as part

Lx Ly Lz   Lx Ly

        X   Y   Z   X   Y   Z'

E   E

(a)   (b)
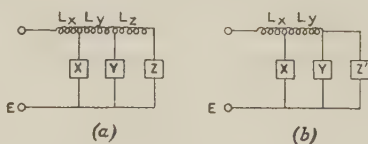
Fig. 15.—Equivalent circuit of the main junction blocks.

of the residual inductance of Z, so that the end two shunt ele-
ments can be assumed to be in parallel, as in Fig. 15(b), when Z
is modified to Z'. It will thus be seen that there are only three
possible equivalent circuits for the junction. Had the trimmer
capacitors as well as the inductor been connected individually
to the junction block, and not connected in parallel within a
second cavity and joined to the junction block by a common
line, there would have been five shunt elements and the number
of possible equivalent circuits of the junction would have been
60. With the arrangement adopted it is possible to identify the
appropriate equivalent circuit by simple qualitative experiments.

In order to illustrate the erroneous results which can be
obtained by not selecting the correct equivalent circuit it was
found that, if L_j was taken as zero and the experimental results
fitted to this circuit, L_k was determined as +324 pH. Using the
same experimental data and fitting it to the true equivalent
circuit, L_j becomes -324 pH and L_k becomes -276 pH—a
difference of 924 pH in the apparent total equivalent junction
inductance, which is sufficient to cause an error of several per cent
at 200 Mc/s in the measurement of a large admittance.

The negative values obtained for L_j and L_k are caused by the
mutual inductance between the various circuit element connections
within the main cavities predominating over the common
inductance due to the finite area of the junctions.

(12) STANDARD RESISTOR G_s

Any shunt capacitance across G_s does not affect the accuracy
of conductance measurement across the B-arm but it does
introduce a second-order effect for the A-arm measurements,
because, in this instance, the bridge equation contains the
effective series resistance of G_s [eqns. (5) and (7)]. In addition,
a first-order effect is caused in capacitance measurement for
both arms of the bridge. If the effective shunt capacitance of
G_s is C_{sg}, eqns. (4) and (6) are modified as follows:

Unknown across the B-*arm*

$$C_{xb} = \Delta C_b + \frac{G_{xb}}{G_s}C_{sg} . \quad . \quad . \quad . \quad (35)$$

Unknown across the A-*arm*

$$C_{xa} = \Delta C_a - \frac{G_{xa}}{G_s}C_{sg} \quad . \quad . \quad . \quad (36)$$

It will be seen that C_{sg} produces equal but opposite effects for
the two methods of measurement. This is generally true for all
effects associated with the residuals of the series elements, and
is a characteristic of this bridge circuit which makes the dual
feature a possibility. The value of C_{sg} can be determined by
measuring the capacitance of a resistor across both arms at a
low frequency where the d.c. values of G_s and G_x are effective.
For this condition eqns. (35) and (36) can be equated to one
another as follows:

$$C_{sg} = -\frac{G_s}{2G_x}(\Delta C_{bB} - \Delta C_{aA}) . \quad . \quad . \quad (37)$$

where ΔC_{bB} is the change in C_b for the B-arm connection and
ΔC_{aA} is the change in C_a for the A-arm connexion. The experi-
mental value obtained for C_{sg} was as follows:

$$C_{sg} = -0.068 \pm 0.001 \text{ pF} \quad . \quad . \quad . \quad (38)$$

The dual bridge is also used to establish the effective con-
ductance of G_s in terms of the susceptance standards, whose
values are calculated from length and time measurements. In a
similar manner to that adopted for determining C_{sg} by measuring
the same resistor across both arms of the bridge, eqns. (3) and (5)

can be equated to one another. When this is done, and using relationship (7), we get

$$G_s = \frac{\omega C_1 C_2}{C_3} \sqrt{\frac{\Delta C_{bA}}{\Delta C_{aB}}} \frac{1}{\sqrt{\left(1 + \frac{\omega^2 C_{sg}^2}{G_s^2}\right)}} \quad . \quad . \quad (39)$$

where ΔC_{bA} and ΔC_{aB} refer to the A- and B-arm methods of connection, respectively.

In the preceding Sections the methods have been described whereby the values of $C_1 C_2 / C_3$, ΔC_a and ΔC_b are referred back to the susceptance standards. It has also been shown how frequency errors associated with $C_1 C_2 / C_3$ and the common earth impedances are eliminated. In the evaluation of the last square-root term in eqn. (39) it is sufficient to use the d.c. conductance of G_s, because the value of this term is very small, and amounts to only a few parts in 10^4 at 200 Mc/s. When the experiments to determine G_s are carried out at high frequencies, small corrections have to be applied for the effect of the junction inductances L_j and L_k. The value of G_s was measured at various frequencies up to 250 Mc/s, and it was found that the change from the d.c. value was very small, and amounted to only 0.12% at 200 Mc/s. The estimated uncertainty in the determinations was 0.1% at all frequencies. These relationships are expressed as $G_s = G_{ss}(1 + z)$, where G_{ss} is the d.c. conductance.

(13) FINAL 1-BAR BALANCE DIFFERENCE EQUATIONS

Eqn. (34) relates admittance changes in the 3-bar values of the shunt admittances of the A- and B-arms. In order that these values may be expressed in terms of 1-bar admittance changes it is necessary to transform them through the junction impedances together with the 2-bar values relating to C_a and C_b (Fig. 3). When this is done we get the final 1-bar balance difference equation for the unknown admittance $\bar{Y}_x = \bar{G}_x + j\omega \bar{C}_x$:

Unknown across the B-arm

$$\bar{G}_{xb} = \frac{G_{ss}(1+z)C_3}{C_1 C_2}\Delta\bar{C}_a[1 + \omega^2 L_j(2\bar{C}_{b2} - \bar{C}_{a1} - \bar{C}_{a2} - C_5)$$
$$- \omega^2 L_k(\Delta\bar{C}_a + C_5)] + \omega^2 R_{0b}(\bar{C}_{b1}^2 - \bar{C}_{b2}^2) \quad . \quad (40)$$

$$\bar{C}_{xb} = \Delta\bar{C}_b + \frac{\bar{G}_{xb}}{G_s}C_{sg} + (L_j + L_k)\bar{G}_{xb}^2$$
$$- \frac{R_{0a}G_sC_3}{C_1 C_2}(\bar{C}_{a1}^2 - \bar{C}_{a2}^2) + C_f \quad . \quad (41)$$

where
$$C_5 = 2\left(C_1 + C_2 + \frac{C_1 C_2}{C_3} - C_3\right)$$
$$= \frac{2C_1 C_2}{C_3} \text{ when } C_1 = C_2 = \tfrac{1}{2}C_3$$
$$= 5\,\text{pF}$$

It should be noted that the magnitude of the $\omega^2 L_j$ correction term is dependent upon the difference between twice the final value of $\bar{C}_b$ and the sum of the initial and final values of $\bar{C}_a$, i.e. $2\bar{C}_{b2} - (\bar{C}_{a1} + \bar{C}_{a2})$. Because C_b is approximately twice C_a and because the former is normally set to its mid-position and the latter to its maximum value for initial balance, the capacitance in this correction term is quite small. Also the sign of the $\omega^2 L_k$ term is the opposite of the $\omega^2 L_j$ term, which also reduces the magnitude of the overall correction. The value of this correction in the main bracket is quite small at 200 Mc/s and rarely exceeds

1%; for most measurements it amounts to only a few parts in 10^3. The final term arises from the resistance of C_b, and this is usually fairly small, but it can be as large as 40 micromhos for a large susceptance change at 200 Mc/s. As already mentioned the utilization of tellurium-copper instead of brass for the electrodes of the main capacitors would reduce this correction to one-half. Reference has already been made to the correction $(1 + z)$, which is also very small and is equal to 1.0012 at 200 Mc/s. All the correction terms in eqn. (41) are independent of frequency except R_{0a}, which increases as its square root. The first, second and third correction terms containing C_{sg}, $(L_j + L_k)$ and R_{0a} are maxima when $\bar{G}_{xb}$ is greatest. For a value of 50 millimhos these corrections amount to $-0.68\,\text{pF}$, $-1.5\,\text{pF}$ and $-0.05\,\text{pF}$, respectively. The values of $\Delta\bar{C}_a$ and $\Delta\bar{C}_b$ are, of course, calculated from the formulae embodying the residual inductances of these elements. The corrections in this case are large, and amount to approximately 10% and 15%, respectively, at 200 Mc/s. Because the equivalent circuits are so precisely defined the values are given to within 0.1% of the susceptance standards, which are themselves within 0.1% of the absolute values.

The complementary 1-bar balance difference equations for the A-arm method of connection are as follows:

$$\bar{G}_{xa} = \omega^2 R_s \frac{C_1 C_2}{C_3}\Delta\bar{C}_b[1 + \omega^2 L(2\bar{C}_{a2} - \bar{C}_{b1} - \bar{C}_{b2} + C_5)$$
$$+ \omega^2 L_k(\Delta\bar{C}_b + C_5)] + \omega^2 R_{0a}(\bar{C}_{a1}^2 - \bar{C}_{a2}^2) \quad . \quad (42)$$

$$\bar{C}_{xa} = \Delta\bar{C}_a - \frac{\bar{G}_{xa}}{G_s}C_{sg} + (L + L_k)\bar{G}_{xa}^2$$
$$- \frac{\omega^2 R_{0b}C_1 C_2}{G_s C_3}(\bar{C}_{b2}^2 - \bar{C}_{b1}^2) + C_f \quad . \quad (43)$$

The values of the components of the unknown admittance $Y_x = G_x + j\omega C_x$ in the plane of measurement are obtained by substituting the above 1-bar values in eqns. (11) and (12).

In the practical application of corrections a considerable amount of the data can be obtained from curves. Furthermore, the labour involved can be reduced considerably by employing a Table for computing the corrections in conjunction with an electrically operated calculating machine. For example, a set of six measurements at one frequency occupy approximately two hours for the application of full corrections. This compares very favourably with other methods such as a slotted line, where the application of corrections for non-uniformity of probe penetration, probe reflection, the resistance of the line, skin effect, etc., would take considerably longer and even then would not yield the same precision as the bridge method described.

(14) VERIFICATION OF BRIDGE ACCURACY

The various processes have been described whereby the calibration of the bridge is effected in terms of the fixed susceptance standards. Having reached this stage, it does not necessarily mean that the bridge will give the correct result when measuring any complex admittance within its range. This condition will only obtain when the correct equivalent circuit has been found. The importance of choosing the correct equivalent circuit was emphasized very forcibly at one stage in the verification of bridge accuracy when the incorrect equivalent circuit of the main junction blocks had been used. Although the bridge was quite satisfactory for the independent measurement of substantially pure susceptances and conductances, relatively large errors arose when measuring complex admittances. The cause of the anomalous results was not known at the time, and some

18 months' additional experimental work and circuit analysis was necessary before the anomaly was resolved.

The method employed to verify the equivalent circuit is based on the well-known property of a transmission line, which has a complex transformation ratio between its ends. If such a line is terminated in a complex admittance, each component of the input admittance will be dependent upon both components of the terminating admittance in addition to the line parameters. The longest standard coaxial capacitor was used for the line and a range of coaxial resistor mounts, having conductances between 5 and 40 millimhos, were used as terminating admittances. The resistor mounts were first measured by means of the bridge at a frequency of 200 Mc/s and then mounted in turn on the end of the capacitor line. The complex admittances so formed were then measured by means of the bridge, and the values obtained were compared with the calculated values. From the difference between these two sets of results it was possible to estimate the absolute accuracy of the bridge. A typical set is given in Table 1.

Table 1

COMPLEX ADMITTANCE STANDARDS AT 200 Mc/s

Conductance component					
	mmhos	mmhos	mmhos	mmhos	mmhos
Calculated	8·562	16·618	21·240	28·844	38·045
Measured ..	8·566	16·596	21·179	28·786	38·056
Ratio ..	0·999 5	1·001 3	1·002 9	1·002 0	0·999 7

Capacitance component					
	pF	pF	pF	pF	pF
Calculated	27·301	25·093	22·923	17·723	0·889
Measured ..	27·324	25·120	22·916	17·703	0·924
Difference..	+0·023	+0·027	−0·007	−0·020	+0·035

Approximate phase angle of complex admittance				
76°	62°	54°	38°	2°

Nominal value of terminating resistor				
mmhos	mmhos	mmhos	mmhos	mmhos
5	10	13·33	20	40

The difference between the two sets of results is affected by the uncertainty in two sets of measurements, in addition to the uncertainty in the calculated parameters of the capacitor line. Examination of these results, taking into account the uncertainties in the residual parameters, confirms that the inaccuracy of the bridge at 200 Mc/s is as follows:

Conductance... 0·2% ± 5 micromhos.
Capacitance... 0·2% ± 0·005 pF for conductances up to 10 mmhos
                    ± 0·02 pF for conductances up to 30 mmhos
                    ± 0·05 pF for conductances up to 50 mmhos

Other measurements carried out at lower frequencies confirm that the inaccuracy up to 100 Mc/s is as follows:

Conductance... 0·1% ±2 micromhos
Capacitance... 0·1% ±0·005 pF for conductances up to 10 mmhos
                    ±0·02 pF for conductances up to 30 mmhos
                    ±0·05 pF for conductances up to 50 mmhos

A diagram is given in Fig. 16 which shows the various processes used to calibrate the bridge in terms of length and time

measurements and the ultimate verification of its accuracy in the manner described above.

(15) ANCILLARY EQUIPMENT

The zero transmission condition of the bridge is critically dependent on frequency, so that a very stable c.w. source is required. A crystal-controlled signal source has been developed for this purpose. This instrument provides approximately 35 crystal-controlled frequencies between 5 and 300 Mc/s harmonically related to 5 Mc/s. The source e.m.f. and impedance are 0·5 volt and 75 ohms, respectively, and the screening is as good as that associated with a first-grade signal generator. This latter requirement is necessary because any leakage from the source or the interconnecting cable to the bridge can cause undesirable hand-capacitance effects. Pyrotenax or lead-covered coaxial cables have been found satisfactory in this respect.

A good communication receiver is a satisfactory detector for the bridge. In order to overcome the difficulty of local-oscillator drift in the receiver at very high frequencies, an amplitude/frequency-modulated crystal mixer is used between the bridge and the receiver. The local oscillator consists of a conventional signal generator, amplitude-modulated but also producing frequency modulation. With this arrangement the instrument behaves like a normal audio-frequency bridge up to 300 Mc/s with a discrimination and precision of setting of at least two parts in 10^5 at initial balance and better than one part in 10^4 at final balance when measuring an admittance of $50 \pm j50$ millimhos.

(16) MEASUREMENT TECHNIQUE

In order to realize the full precision of the bridge great care must be exercised in the method of connection of the unknown to the terminals. If the unknown carries a conventional type of plug, the most satisfactory method of connection is by means of a cone having the same characteristic impedance as the plug attached to the unknown and with an outer diameter at the bridge end the same as that of the bridge terminals. A sectional view of such a connector is given in Fig. 17.

When the cone impedance is greater than 24·3 ohms this gives rise to a step in the inner conductor at the bridge terminals. This can be represented by a capacitance C_d concentrated in the plane of the junction. The curve in this Figure gives the value of C_d for various cone impedances. This curve is based on a cylindrical system and was prepared from data given in Reference 5. There will be no significant difference if the semi-angle of the outer cone is small, i.e. 4° or less. A suitable length of cone for most applications is 2·254 in, which makes $\beta l = 0·120 0$ rad at 100 Mc/s if the velocity is assumed to be $2·997 1 \times 10^{10}$ cm/s.

It will be seen that the cone is taken as part of the unknown for the purpose of measurement. Initial balance is obtained with nothing attached to the bridge terminals, the cone inner is then screwed in, the cone outer attached by means of the three screws and the unknown then plugged into the cone. If the admittance of the unknown in the plane B is Y_r, and this admittance transformed through the cone to the plane A is Y_i, the bridge will measure $Y_i + j\omega C_d$:

i.e.
$$G_i = G_{xb} \qquad . \quad . \quad . \quad . \quad . \quad (44)$$

and
$$C_i = C_{xb} - C_d \qquad . \quad . \quad . \quad . \quad (45)$$

Y_r, in terms of Y_i and the cone parameters, can now be calculated from the usual line equation. It has been found convenient to use the following formulae for this purpose because it is not
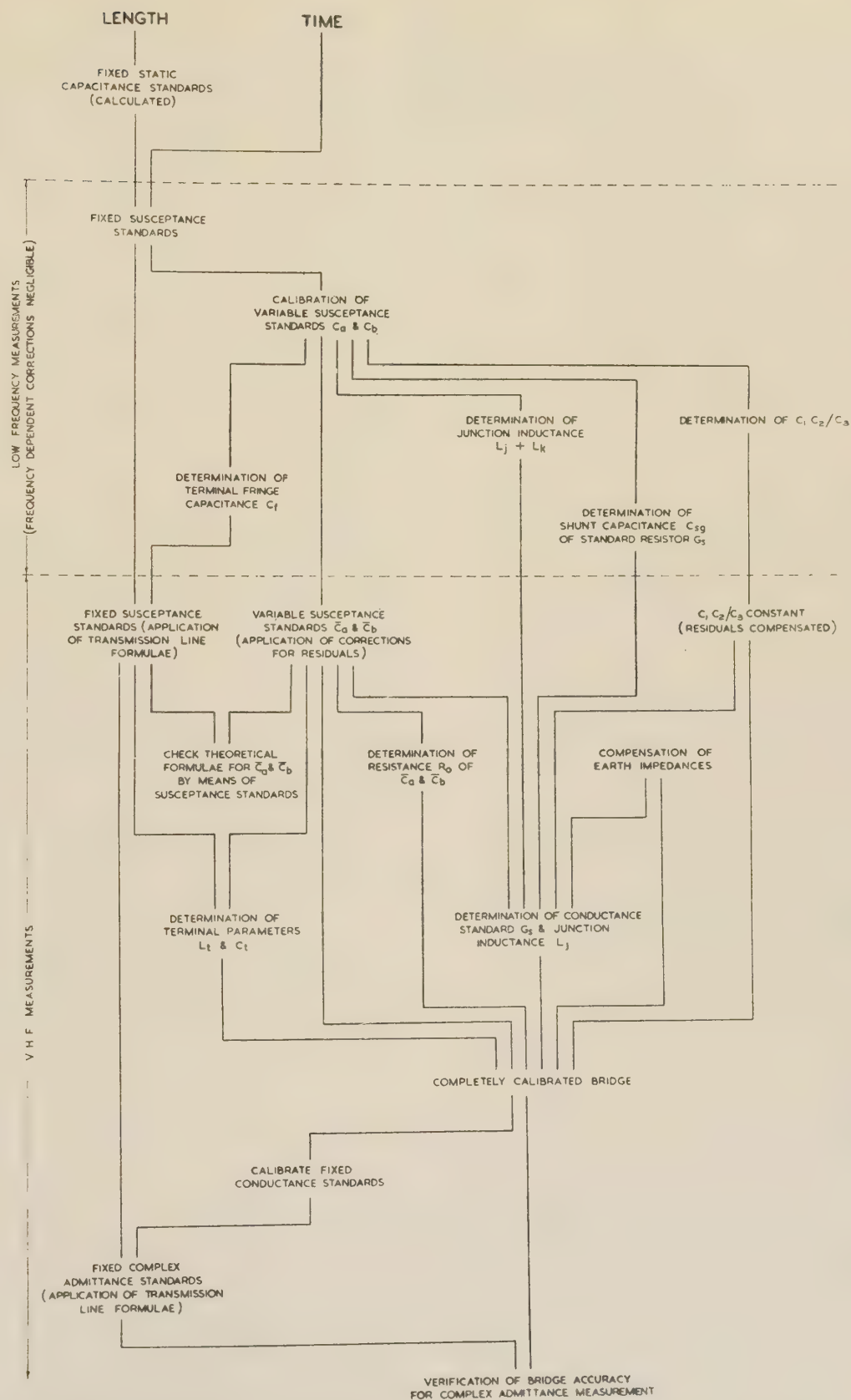
Fig. 16.—Basic method for the standardization of the bridge by means of length and time measurements.
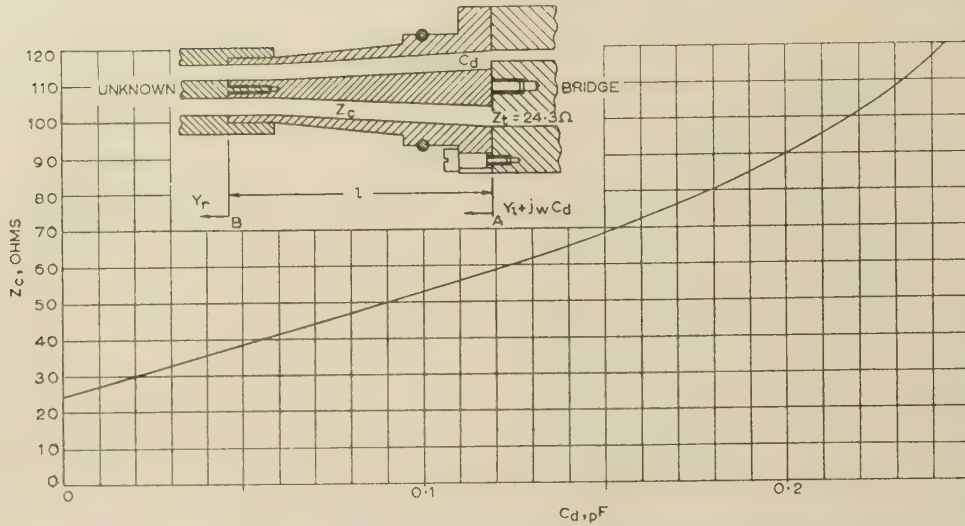
Fig. 17.—Discontinuity capacitance between the bridge terminals and a line of characteristic impedance Z_c, having the same outer diameter.

necessary to convert capacitances into susceptances, and also all the factors are dimensionless:

$$G_r = \frac{G_i(1 + \tan^2 \beta l)}{\left(1 + \dfrac{C_i}{C'} \tan^2 \beta l\right)^2 + G_i^2 Z_c^2 \tan^2 \beta l} \qquad . \quad . \quad (46)$$

$$C_r = \frac{C'\left[\left(\dfrac{C_i}{C'} - 1\right)\left(1 + \dfrac{C_i}{C'} \tan^2 \beta l\right) + G_i^2 Z_c^2\right]}{\left(1 + \dfrac{C_i}{C'} \tan^2 \beta l\right)^2 + G_i^2 Z_c^2 \tan^2 \beta l} \qquad . \quad (47)$$

where $Z_c = Z_\infty[1 + \zeta(1 + \gamma)]$

Z_∞ is calculated from the dimensions of the cone.

i.e. $Z_\infty = 138 \log_{10} r_0/ri$.

$(1 + \zeta)$ is given by curves similar to Fig. 9, but for the appropriate characteristic impedance and for a cylindrical system having an outer diameter of $0 \cdot 750$ in.

$(1 + \gamma)$ is given by Fig. 18.

$C' = C \dfrac{\tan \beta l}{\beta l}$

$C = l/v_\infty Z_\infty$.

$\beta = \omega/v$.

$v = v_\infty/[1 + \zeta(1 + \gamma)]$.

$v_\infty = 2 \cdot 997 1 \times 10^{10}$ cm/s.

$l =$ length of cone, cm.

No approximations have been made in the derivation of these formulae from the transmission-line equation except for the assumption that the series resistance and shunt conductance of the cone are both negligibly small. It should also be remembered that the formula for Z_∞ is for a cylindrical system, so that it will be an approximation for a conical system; in addition the axial length l is an approximation on the slant length of the cone. Again, if the semi-angle of the cone outer is not greater than 4° these differences will not be significant.

The factor $(1 + \gamma)$ arises because the velocity in a conical system is not uniform. This effect is also produced by current penetration, which causes the inductance per unit length to

increase as the cone diameter becomes smaller. This factor when multiplied by ζ gives the average reciprocal velocity factor $[1 + \zeta(1 + \gamma)]$. It can be shown that $(1 + \gamma)$ is independent of frequency and the characteristic impedance and length of the cone,

i.e. $$(1 + \gamma) = \frac{2 \cdot 303 \log_{10} r_0/r_0'}{1 - r_0'/r_0} \qquad . \quad . \quad (48)$$

Even though $(1 + \gamma)$ is independent of frequency the correction will not be significant at the higher frequencies because ζ will then be small. The factor is plotted in Fig. 18.
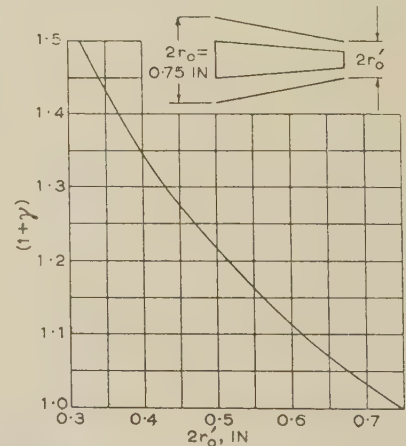
Fig. 18.—Correction factor $(1 + \gamma)$ for a conical coaxial line of any characteristic impedance, having a maximum outer diameter of $0 \cdot 750$ in.

Eqns. (46) and (47) give the components of the unknown admittance in the plane B. In some cases it is not possible to achieve a negligible mismatch in this plane owing to the poor design of the plug attached to the unknown. Then there is very little that can be done, which means that the admittance cannot be defined so precisely.

This difficulty can be overcome, and at the same time the use of cones avoided, by employing the universal coaxial connector system to which reference has already been made.[3]

(17) CONCLUSIONS

A bridge has been designed having a range of measurement associated with unbalanced transmission lines and an accuracy which is sufficient to enable an improvement to be made of at least one order in accuracy in the design of apparatus and measuring instruments for the basic electrical quantities at frequencies up to several hundred megacycles per second. The instrument incorporates a number of features whereby the effects of unavoidable residual parameters, associated with the common earth impedances and the three series capacitors, are substantially compensated, thus enabling the true r.f. conductance of the bridge resistor to be established in terms of the fundamental units to an order of precision which has hitherto not been possible. Compensation of the small residual correction term containing C_{sg}, in the formula for the unknown capacitance, could readily be accomplished by means of a small preset capacitor in shunt with the bridge resistor but external to the resistor mount. This, together with the other methods referred to, would then provide compensation for all the residuals of the series elements except the frequency-dependent parameter of the resistive component of the bridge resistor, although this has very nearly been achieved by the design of coaxial mount adopted.

The negative values obtained for the junction inductances show that the mutual impedances between various parts of the junction blocks and the circuit elements attached thereto are greater than the positive common impedances caused by the finite area of the junctions. This suggests that it should be possible to realize a practical design in which these two effects cancel, thereby causing the 1-bar, 2-bar and 3-bar points to coalesce and reducing to zero all the correction terms containing L_j and L_k.

A new form of variable air capacitor has been developed in which the residual inductance is caused to vary in such a manner that the behaviour at high frequencies can be utilized for the purpose of measuring the residual parameters very accurately. The formulae developed for this capacitor contain no approximations, in contrast to the well-known method[6] for determining the inductance of a capacitor of conventional design in which this parameter is assumed to be independent of capacitance setting. The error arising from the formula for the effective capacitance at frequencies up to 200 Mc/s is no greater than that due to the uncertainty in the calibration at 1 kc/s, namely 0·1%.

It is considered that, by utilizing similar methods for the compensation of residual effects, it should be possible to construct a bridge of comparable accuracy for operation up to approximately 500 Mc/s.

A range of admittance standards has also been developed and calibrated by means of the bridge. These are suitable for use as transfers between laboratories and for assessing the accuracy of other admittance-measuring instruments fitted with suitable terminals. Attention has been drawn to the need for exercising care in the design of the terminations of standards and immittance-measuring instruments generally at very high frequencies. This is a matter upon which universal agreement must be reached before the standardization of immittance at these frequencies can be considered as placed on a firm foundation. With this end in view, a universal coaxial connector system[3] has been designed, which is considered suitable for precision radio-frequency measuring instruments of all types for use at frequencies up to 4 000 Mc/s. With the general adoption of such a system the problems associated with the standardization and measurement of other electrical quantities would also be much simplified. This would ultimately lead to better accuracy in measurement generally, with consequent improvement in design performance, not only of the measuring instruments, but also of the operational equipment.

(18) REFERENCES

(1) HARRIS, I. A.: 'The Theory and Design of Coaxial Resistor Mounts for the Frequency Band 0–4 000 Mc/s', *Proceedings I.E.E.*, Monograph No. 132 R, May, 1955 (**103** C, p. 1).

(2) WOODS, D.: 'A Precision V.H.F. Twin-T Dual Admittance Bridge and Associated Standards', Unpublished Ministry of Supply Report, No. A.I.D. Rad.Elect./7, Parts I and II; January, 1950 (available through the Central Radio Bureau).

(3) WOODS, D.: 'Co-axial Terminations and a Connector System for Precision Radio Frequency Measuring Instruments', Unpublished Ministry of Supply Report, No. A.I.D. Rad. Elect./12; April, 1952 (available through the Central Radio Bureau).

(4) SINCLAIR, D. B.: 'The Twin-T. A New Type of Null Instrument for Measuring Impedance at Frequencies up to 30 Megacycles', *Proceedings of the Institute of Radio Engineers*, 1940, **28**, p. 310.

(5) WHINNERY, J. R., JAMIESON, H. W., and ROBBINS, T. E.: 'Coaxial Line Discontinuities', *ibid.*, 1944, **32**, p. 695.

(6) FIELD, R. F., and SINCLAIR, D. B.: 'A Method for Determining the Residual Inductance and Resistance of a Variable Air Condenser at Radio Frequencies', *ibid.*, 1936, **24**, p. 255.

(7) 'Apparatus for Measuring Electrical Admittance', British Patent No. 681927, October, 1952.

621.316.5.015.3

ANALYSIS OF TRANSIENT CONDITIONS ACCOMPANYING SWITCHING OPERATIONS IN THREE-PHASE POWER SYSTEMS

J. H. BANKS, M.Sc., Associate Member.

SUMMARY

The paper describes matrix methods whereby the transient voltages and currents produced during switching operations in 3-phase power systems may be evaluated, taking into account appreciable mutual impedances existing between phases of the various items of plant. Operational voltage and impedance matrices are derived which define the transient performance of a network during the clearance of a shunt fault, the individual phases of the circuit-breaker being considered to clear sequentially. The successive constraints imposed upon the network currents are defined by transformation matrices using methods based on the work of Kron. The influence of phase mutual impedances upon the transient over-voltages produced during the isolation of single-phase arcing faults on double-circuit overhead lines is also considered. A simplified mathematical treatment of the effect of circuit-breaker arc restrikes upon the transient response of a power system is described in a concluding Section dealing with resistance switching of overhead-line charging current.

LIST OF PRINCIPAL SYMBOLS

$[\bar{V}(p)], [\bar{I}(p)], [Z(p)]$ = Voltage, current and impedance operational matrices, respectively, of the system in terms of phase or symmetrical component parameters.

$[C]$ = Connection matrix.

$[\bar{C}_t]$ = Complex conjugate of the transpose of C.

1, 2, 0 = Subscripts denoting positive, negative and zero phase sequence components, respectively.

$h = -0·5 + j0·866.$

(1) INTRODUCTION

When unbalanced faults occur on 3-phase power systems, the transient voltages and currents are not sinusoidal, and may contain pronounced harmonics; the magnitudes of these quantities are often an important consideration when investigating the stressing of the insulation of switchgear, transformers, etc., under transient conditions.

Mutual impedances existing between phases of transmission lines, cables, and synchronous machines in particular, may have an appreciable effect upon the magnitudes of transient voltages and currents produced during switching operations in the power system. Furthermore, the presence of simultaneous dissymmetries imposes further constraints which may increase the severity of the transient disturbance.

The recent analysis of the performance of synchronous generators by Ching and Adkins[1] contained a rigorous statement of the equations of performance of the 3-phase synchronous machine expressed in terms of direct- and quadrature-axis quantities under both balanced and unbalanced short-circuit conditions. The solution for the balanced short-circuit was

expressed in the form of linear differential equations with constant coefficients in which the applied voltages were known. In the unbalanced case, the analysis was derived in the form of α, B, 0 components, the resultant expressions being linear differential equations with variable coefficients. The equations of performance of an interconnected power system may therefore be defined taking into account all appreciable mutual impedances of the various items of plant comprising the system, and the transient conditions produced by switching operations at any point in the system subsequently evaluated.

Typical switching conditions involving simultaneous dissymmetries are considered in the present paper, and an operational matrix treatment of these is described, using methods based on the work of Kron.[2] The methods adopted enable the voltage and impedance matrices for the various unbalanced conditions to be derived from those pertaining to an initial balanced (primitive) condition. In order to transform the matrices defining the primitive condition to those representing the unsymmetrical states, connection matrices must be deduced which express the relationships between the parameters of the primitive condition and those for the succeeding unsymmetrical conditions.

During the operation of a 3-phase circuit-breaker, the sequential clearance of the main contacts imposes successive unsymmetrical constraints upon the currents flowing in the network. This condition is investigated in Section 3, taking into account the major mutual impedances existing between phases of the transmission line and synchronous machine involved. The analysis has been performed using the symmetrical components of Fortescue, but the use of other systems of components can effect further simplification.

The isolation of persistent single-phase arcing faults on double-circuit overhead lines produces transient over-voltages which may be aggravated by the influence of capacitance coupling from the other phases. A matrix analysis of the sequential dissymmetries involved during fault clearance is described in Section 4.

The occurrence of discontinuities such as circuit-breaker arc restrikes exerts a considerable influence upon the shape of the transient recovery voltage characteristic. Whilst the mathematical definition of phenomena of this nature may not be stated precisely at the commencement of an analysis, due to the influence of random physical disturbances such as oil turbulation, it is, however, possible to calculate the transient voltage response of the network for a range of boundary conditions, and consequently to deduce the most onerous case. The influence of arc restriking is illustrated by considering the transient conditions accompanying the interruption of transmission-line charging current when resistance switching is employed.

The extensive developments in programming techniques for automatic digital-computing machines have included the preparation of standard computer sub-routines for the solution of simultaneous differential equations.[3] Full advantage may therefore be taken of these and allied standard programmes when presenting transient problems of this nature to computing machines for solution.

(2) BASIC TECHNIQUES

(2.1) Operational Matrices of the Three-Phase Synchronous Machine

The reduced operational voltage equations of the three armature circuits of a 3-phase synchronous machine after eliminating the field and damper current equations have been derived by Ching and Adkins[1] and are stated below.

$$[\bar{v}(p)] = [Z(p)] \cdot [\bar{i}(p)] \quad . \quad . \quad . \quad . \quad (1)$$

where

$$[\bar{v}(p)] = \begin{array}{c} a \\ b \\ c \end{array} \begin{array}{|c|} \hline \bar{v}_{at}(p) + \bar{e}_a(p) - \tfrac{1}{3}p\{\varepsilon^{-j2\lambda}\overline{L}_D(p+j\omega)[\bar{i}_a(p) + j2\omega + h\bar{i}_b(p+j2\omega) + h^2\bar{i}_c(p+j2\omega)] + \text{conj}\} \\ \hline \bar{v}_{bt}(p) + \bar{e}_b(p) - \tfrac{1}{3}p\{h\varepsilon^{-j2\lambda}\overline{L}_D(p+j\omega)[\bar{i}_a(p+j2\omega) + h\bar{i}_b(p+j2\omega) + h^2\bar{i}_c(p+j2\omega)] + \text{conj}\} \\ \hline \bar{v}_{ct}(p) + \bar{e}_c(p) - \tfrac{1}{3}p\{h^2\varepsilon^{-j2\lambda}\overline{L}_D(p+j\omega)[\bar{i}_a(p+j2\omega) + h\bar{i}_b(p+j2\omega) + h^2\bar{i}_c(p+j2\omega)] + \text{conj}\} \\ \hline \end{array} \quad . \quad . \quad . \quad (2)$$

$$[Z(p)] = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{|c|c|c|} \multicolumn{1}{c}{a} & \multicolumn{1}{c}{b} & \multicolumn{1}{c}{c} \\ \hline r_a + \tfrac{1}{3}p[L_0 + \overline{L}_S(p+j\omega) + \text{conj}] & \tfrac{1}{3}p[L_0 + h^2\overline{L}_S(p+j\omega) + \text{conj}] & \tfrac{1}{3}p[L_0 + h\overline{L}_S(p+j\omega) + \text{conj}] \\ \hline \tfrac{1}{3}p[L_0 + h\overline{L}_S(p+j\omega) + \text{conj}] & r_a + \tfrac{1}{3}p[L_0 + \overline{L}_S(p+j\omega) + \text{conj}] & \tfrac{1}{3}p[L_0 + h^2\overline{L}_S(p+j\omega) + \text{conj}] \\ \hline \tfrac{1}{3}p[L_0 + h^2\overline{L}_S(p+j\omega) + \text{conj}] & \tfrac{1}{3}p[L_0 + h\overline{L}_S(p+j\omega) + \text{conj}] & r_a + \tfrac{1}{3}p[L_0 + \overline{L}_S(p+j\omega) + \text{conj}] \\ \hline \end{array} \quad (3)$$

$$\bar{i}(p) = \begin{array}{|c|} \hline \bar{i}_a(p) \\ \hline \bar{i}_b(p) \\ \hline \bar{i}_c(p) \\ \hline \end{array} \quad . \quad . \quad . \quad . \quad (4)$$

The symbols r, L and M denote, respectively, resistance, self-inductance and mutual inductance, and the suffixes indicate the windings to which they relate. A full definition of the elements of the voltage and impedance matrices (2) and (3) is given in Reference 1. The term 'conj' is used to indicate the conjugate function obtained by replacing j by $-j$. It is not the same as the numerical conjugate when p is complex.

The corresponding symmetrical-component operational-voltage matrix equation can be evaluated from eqn. (1) by performing the well-known transformations given below in eqns. (5) and (6).

$$[\bar{v}_s(p)] = [T_s]^{-1} \cdot [\bar{v}(p)] \quad . \quad . \quad . \quad . \quad (5)$$

and

$$[Z_s(p)] = [T_s]^{-1} \cdot [Z(p)] \cdot [T_s] \quad . \quad . \quad . \quad (6)$$

where $[T_s]^{-1} = $

$\tfrac{1}{3}$	$\tfrac{1}{3}h$	$\tfrac{1}{3}h^2$
$\tfrac{1}{3}$	$\tfrac{1}{3}h^2$	$\tfrac{1}{3}h$
$\tfrac{1}{3}$	$\tfrac{1}{3}$	$\tfrac{1}{3}$

$\quad . \quad . \quad . \quad (7)$

and $[T_s] = $

1	1	1
h^2	h	1
h	h^2	1

$\quad . \quad . \quad . \quad (8)$

The symmetrical-component voltage equation is therefore

$$[\bar{v}_s(p)] = [Z_s(p)] \cdot [\bar{i}_s(p)] \quad . \quad . \quad . \quad (9)$$

For the case of the 3-phase synchronous machine,

$$[\bar{v}_s(p)_M] = \begin{array}{|c|} \hline \bar{v}_1(p)_M \\ \hline \bar{v}_2(p)_M \\ \hline \bar{v}_0(p)_M \\ \hline \end{array} \equiv \begin{array}{|c|} \hline \tfrac{1}{3}[\bar{v}_a(p) + h\bar{v}_b(p) + h^2\bar{v}_c(p)] \\ \hline \tfrac{1}{3}[\bar{v}_a(p) + h^2\bar{v}_b(p) + h\bar{v}_c(p)] \\ \hline \tfrac{1}{3}[\bar{v}_a(p) + \bar{v}_b(p) + \bar{v}_c(p)] \\ \hline \end{array} \quad (10)$$

and $[Z_s(p)_M] = $

Z_{11M}	Z_{12M}	Z_{10M}
Z_{21M}	Z_{22M}	Z_{20M}
Z_{01M}	Z_{02M}	Z_{00M}

$\quad . \quad . \quad (11)$

where

$$Z_{11M} = \tfrac{1}{3}(Z_{aa} + Z_{bb} + Z_{cc}) - \tfrac{1}{3}(Z_{ab} + Z_{bc} + Z_{ac})$$
$$Z_{12M} = \tfrac{1}{3}(Z_{aa} + h^2Z_{bb} + hZ_{cc}) + \tfrac{2}{3}(hZ_{ab} + h^2Z_{ac} + Z_{bc}), \text{ etc.}$$

(2.2) Short Single-Circuit Three-Phase Transmission Line

(2.2.1) Shunt Capacitive Impedances.

The derivation of the voltage equations for the four line conductors of a 3-phase transmission line with earth wire has been described by Mortlock and Humphrey Davies.[4] Writing these equations in operational form,

$$\left. \begin{array}{l} V_a = Z_{aa}I_a + Z_{ab}I_b + Z_{ac}I_c + Z_{ae}I_e \\ V_b = Z_{ba}I_a + Z_{bb}I_b + Z_{bc}I_c + Z_{be}I_e \\ V_c = Z_{ca}I_a + Z_{cb}I_b + Z_{cc}I_c + Z_{ce}I_e \\ 0 = Z_{ea}I_a + Z_{eb}I_b + Z_{ec}I_c + Z_{ee}I_e \end{array} \right\} \quad . \quad (12)$$

where Z_{aa}, Z_{ab}, etc., are capacitive impedances expressed in operational form $1/(pCmn)$. From these equations, the charging current may be evaluated directly.

Eliminating the equation for the earth wire, the resultant three equivalent equations will be

$$\left. \begin{array}{l} V_a = Z_{a'a'}I_a + Z_{a'b'}I_b + Z_{a'c'}I_c \\ V_b = Z_{b'a'}I_a + Z_{b'b'}I_b + Z_{b'c'}I_c \\ V_c = Z_{c'a'}I_a + Z_{c'b'}I_b + Z_{c'c'}I_c \end{array} \right\} \quad . \quad . \quad (13)$$

If the lumped values of the phase mutual and self-capacitances are now halved, and considered to act at each end of the line X and Y, the equivalent operational 3-phase voltage matrix equation will become

$$
\begin{bmatrix}
\bar{v}_{Xa}(p) \\
\bar{v}_{Xb}(p) \\
\bar{v}_{Xc}(p) \\
\bar{v}_{Ya}(p) \\
\bar{v}_{Yb}(p) \\
\bar{v}_{Yc}(p)
\end{bmatrix}
=
\begin{bmatrix}
Z_{XaXa} & Z_{XaXb} & Z_{XaXc} & & & \\
Z_{XbXa} & Z_{XbXb} & Z_{XbXc} & & & \\
Z_{XcXa} & Z_{XcXb} & Z_{XcXc} & & & \\
& & & Z_{YaYa} & Z_{YaYb} & Z_{YaYc} \\
& & & Z_{YbYa} & Z_{YbYb} & Z_{YbYc} \\
& & & Z_{YcYa} & Z_{YcYb} & Z_{YcYc}
\end{bmatrix}
\cdot
\begin{bmatrix}
\bar{\imath}_{Xa}(p) \\
\bar{\imath}_{Xb}(p) \\
\bar{\imath}_{Xc}(p) \\
\bar{\imath}_{Ya}(p) \\
\bar{\imath}_{Yb}(p) \\
\bar{\imath}_{Yc}(p)
\end{bmatrix}
\quad\ldots\ldots (14)
$$

Eqn. (14) may be transformed into symmetrical component form, thus:

$$
\begin{bmatrix}
\bar{v}_{X1}(p) \\
\bar{v}_{X2}(p) \\
\bar{v}_{X0}(p) \\
\bar{v}_{Y1}(p) \\
\bar{v}_{Y2}(p) \\
\bar{v}_{Y0}(p)
\end{bmatrix}
=
\begin{bmatrix}
Z_{X11} & Z_{X12} & Z_{X10} & & & \\
Z_{X21} & Z_{X22} & Z_{X20} & & & \\
Z_{X01} & Z_{X02} & Z_{X00} & & & \\
& & & Z_{Y11} & Z_{Y12} & Z_{Y10} \\
& & & Z_{Y21} & Z_{Y22} & Z_{Y20} \\
& & & Z_{Y01} & Z_{Y02} & Z_{Y00}
\end{bmatrix}
\cdot
\begin{bmatrix}
\bar{\imath}_{X1}(p) \\
\bar{\imath}_{X2}(p) \\
\bar{\imath}_{X0}(p) \\
\bar{\imath}_{Y1}(p) \\
\bar{\imath}_{Y2}(p) \\
\bar{\imath}_{Y0}(p)
\end{bmatrix}
\quad\ldots\ldots (15)
$$

(2.2.2) Series Impedance.

The phase series-impedance operational matrix equation for a 3-phase overhead line without earth wire is of a form similar to eqn. (13),[4] namely

$$
\begin{bmatrix}
\bar{v}_{aL}(p) \\
\bar{v}_{bL}(p) \\
\bar{v}_{cL}(p)
\end{bmatrix}
=
\begin{bmatrix}
Z_{aLaL} & Z_{aLbL} & Z_{aLcL} \\
Z_{bLaL} & Z_{bLbL} & Z_{bLcL} \\
Z_{cLaL} & Z_{cLbL} & Z_{cLcL}
\end{bmatrix}
\cdot
\begin{bmatrix}
\bar{\imath}_{aL}(p) \\
\bar{\imath}_{bL}(p) \\
\bar{\imath}_{cL}(p)
\end{bmatrix}
\quad (16)
$$

where Z_{aLaL}, Z_{aLbL}, etc., are equivalent series self- and mutual-inductive operational impedances $(R + pL)$.

(2.2.3) Equations of Performance of Short Three-Phase Transmission Line without Earth Wire.

The equivalent 3-phase operational impedance matrix for a short single-circuit 3-phase line without earth wire is a compound diagonal matrix, and the corresponding operational voltage matrix equation is stated below in eqn. (17).

$$
\begin{bmatrix}
\bar{v}_{aL}(p) \\
\bar{v}_{bL}(p) \\
\bar{v}_{cL}(p) \\
\bar{v}_{Xa}(p) \\
\bar{v}_{Xb}(p) \\
\bar{v}_{Xc}(p) \\
\bar{v}_{Ya}(p) \\
\bar{v}_{Yb}(p) \\
\bar{v}_{Yc}(p)
\end{bmatrix}
=
\begin{bmatrix}
Z_{aLaL} & Z_{aLbL} & Z_{aLcL} & & & & & & \\
Z_{bLaL} & Z_{bLbL} & Z_{bLcL} & & & & & & \\
Z_{cLaL} & Z_{cLbL} & Z_{cLcL} & & & & & & \\
& & & Z_{XaXa} & Z_{XaXb} & Z_{XaXc} & & & \\
& & & Z_{XbXa} & Z_{XbXb} & Z_{XbXc} & & & \\
& & & Z_{XcXa} & Z_{XcXb} & Z_{XcXc} & & & \\
& & & & & & Z_{YaYa} & Z_{YaYb} & Z_{YaYc} \\
& & & & & & Z_{YbYa} & Z_{YbYb} & Z_{YbYc} \\
& & & & & & Z_{YcYa} & Z_{YcYb} & Z_{YcYc}
\end{bmatrix}
\cdot
\begin{bmatrix}
\bar{\imath}_{aL}(p) \\
\bar{\imath}_{bL}(p) \\
\bar{\imath}_{cL}(p) \\
\bar{\imath}_{Xa}(p) \\
\bar{\imath}_{Xb}(p) \\
\bar{\imath}_{Xc}(p) \\
\bar{\imath}_{Ya}(p) \\
\bar{\imath}_{Yb}(p) \\
\bar{\imath}_{Yc}(p)
\end{bmatrix}
\quad . (17)
$$

The corresponding symmetrical-component operational voltage matrix equation is formed by transforming the voltage and impedance matrices according to eqns. (5) and (6):

$$
\begin{array}{c}
\bar{v}_{L1}(p) \\
\bar{v}_{L2}(p) \\
\bar{v}_{L0}(p) \\
\bar{v}_{X1}(p) \\
\bar{v}_{X2}(p) \\
\bar{v}_{X0}(p) \\
\bar{v}_{Y1}(p) \\
\bar{v}_{Y2}(p) \\
\bar{v}_{Y0}(p)
\end{array}
=
\begin{array}{ccccccccc}
Z_{L11} & Z_{L12} & Z_{L10} & & & & & & \\
Z_{L21} & Z_{L22} & Z_{L20} & & & & & & \\
Z_{L01} & Z_{L02} & Z_{L00} & & & & & & \\
& & & Z_{X11} & Z_{X12} & Z_{X10} & & & \\
& & & Z_{X21} & Z_{X22} & Z_{X20} & & & \\
& & & Z_{X01} & Z_{X02} & Z_{X00} & & & \\
& & & & & & Z_{Y11} & Z_{Y12} & Z_{Y10} \\
& & & & & & Z_{Y21} & Z_{Y22} & Z_{Y20} \\
& & & & & & Z_{Y01} & Z_{Y02} & Z_{Y00}
\end{array}
\cdot
\begin{array}{c}
\bar{i}_{L1}(p) \\
\bar{i}_{L2}(p) \\
\bar{i}_{L0}(p) \\
\bar{i}_{X1}(p) \\
\bar{i}_{X2}(p) \\
\bar{i}_{X0}(p) \\
\bar{i}_{Y1}(p) \\
\bar{i}_{Y2}(p) \\
\bar{i}_{Y0}(p)
\end{array}
\qquad (18)
$$

(2.3) Kron Transformations

Kron[2] has shown that if the equations of a network $V' = Z'I'$ are known, the equations of another network derived from the first (primitive) by changing the connections, but without increasing the number of independent meshes, may be obtained in the following manner:

The relationship between the mesh currents of the primitive network, I', and those of the new network, I'', is written as

$$I' = CI'' \qquad \ldots \ldots \ldots (19)$$

where C is the 'connection matrix'.

The new voltage matrix V'' is given by

$$V'' = \bar{C}_t V' \qquad \ldots \ldots \ldots (20)$$

and the new impedance matrix Z'' is given by

$$Z'' = \bar{C}_t Z' C \qquad \ldots \ldots \ldots (21)$$

where $\bar{C}_t$ is the complex conjugate of the transpose of C.

(2.4) Stigant's Rule[5]

Stigant's rule enables the impedance matrix of a network to be written down directly by inspection of the network. Independent mesh currents are delineated in each closed mesh, the paths of these currents being chosen to suit the problem under consideration. The elements of the impedance matrix are then derived by the following procedure:

(a) A diagonal element such as Z_{rr} is the total impedance of the path of the rth mesh current.

(b) A non-diagonal element such as Z_{rn} is the impedance common to the rth and nth mesh currents. It is positive if I_r and I_n flow through the impedance in the same direction, and negative if they flow in opposite directions.

(For linear bilateral networks, the impedance matrix is symmetrical and $Z_{rn} = Z_{nr}$.)

As an example of the use of the rule, the equations of the network shown in Fig. 1 are written as

$$
\begin{array}{|c|}
V_y - V_x \\
\hline
V_y
\end{array}
=
\begin{array}{|cc|}
Z_v + Z_x & Z_y \\
\hline
Z_y & Z_y + Z_v
\end{array}
\cdot
\begin{array}{|c|}
I_y \\
\hline
I_x
\end{array}
\qquad (22)
$$

directly from the network.
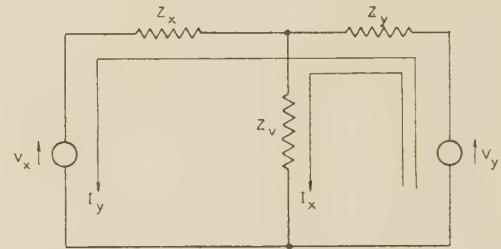
Fig. 1.—Two generators in parallel supplying a load.

(2.5) Symmetrical-Component Operational Matrix Definition of Simple Three-Phase Transmission Network

Consider the case of a 3-phase synchronous alternator feeding a cable network through a short overhead line without earth wire.

Since non-bilateral sequence mutual impedances exist, it is not possible to delineate an equivalent network, but Fig. 2 serves to
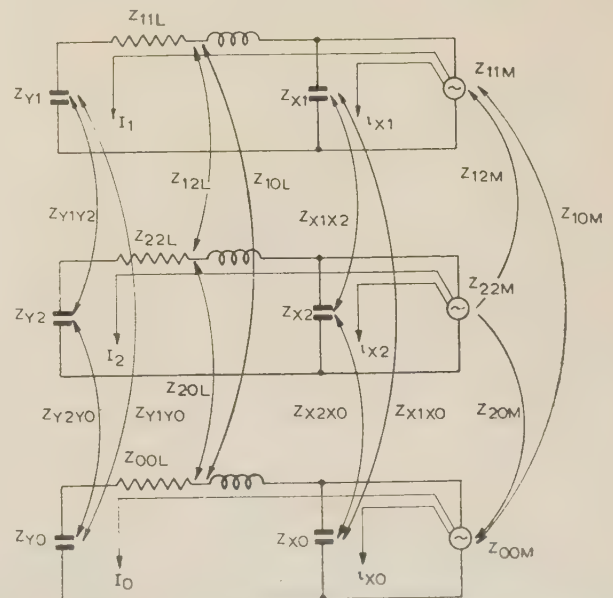
Fig. 2.—Symbolic representation of 3-phase transmission network.

symbolize the conditions applying under normal operating conditions.

The symmetrical-component operational matrix equation may be constructed by applying Stigant's rule in turn to each mesh. The current I_1, for example, on leaving the positive-sequence self-impedance of the synchronous machine Z_{11M} then traverses the positive-sequence series self-impedance of the transmission line Z_{11L} and returns to neutral through the lumped capacitance Z_{Y1}. The operational self-impedance of current I_1 is therefore equal to $Z_{11M} + Z_{11L} + Z_{Y1}$. This current is also linked with current I_2 through three sequence mutual impedances, and therefore the mutual impedance $Z_{12} = Z_{12M} + Z_{12L} + Z_{Y1Y2}$.

The remaining elements of the operational impedance matrix defining the system of Fig. 2 may be derived in a similar manner, and the resultant operational voltage matrix equation is

(3) TRANSIENT RECOVERY VOLTAGE DURING FAULT CLEARANCE

The clearance of persistent faults on 3-phase power systems may cause considerable transient over-voltages as the fault current is being interrupted. It is therefore useful when designing circuit-breakers to have an intimate knowledge of the shape of the recovering voltage transient characteristic for typical circuit conditions during which the circuit-breaker will be called upon to operate.

When considering the clearance of shunt faults, the successive unbalanced conditions existing as the circuit-breaker opens can be defined by current connection matrices. The corresponding unbalanced voltage and impedance matrices are then formed and the transient sequence voltages evaluated. Having transformed these sequence voltages to the phase quantities, the

		1	2	0	X_1	X_2	X_0		
$\bar{V}_1(p)$	1	$Z_{11M} + Z_{11L} + Z_{Y1}$	$Z_{12M} + Z_{12L} + Z_{Y1Y2}$	$Z_{10M} + Z_{10L} + Z_{Y1Y0}$	Z_{11M}	Z_{12M}	Z_{10M}	$\bar{I}_1(p)$	
$\bar{V}_2(p)$	2	$Z_{21M} + Z_{21L} + Z_{Y2Y1}$	$Z_{22M} + Z_{22L} + Z_{Y2}$	$Z_{20M} + Z_{20L} + Z_{Y2Y0}$	Z_{21M}	Z_{22M}	Z_{20M}	$\bar{I}_2(p)$	
$\bar{V}_0(p)$	0	$Z_{01M} + Z_{01L} + Z_{Y0Y1}$	$Z_{02M} + Z_{02L} + Z_{Y0Y2}$	$Z_{00M} + Z_{00L} + Z_{Y0}$	Z_{01M}	Z_{02M}	Z_{00M}	$\bar{I}_0(p)$	(23)
$\bar{v}_{X1}(p)$	X_1	Z_{11M}	Z_{12M}	Z_{10M}	$Z_{11M} + Z_{X1}$	$Z_{12M} + Z_{X1X2}$	$Z_{10M} + Z_{X1X0}$	$\bar{i}_{X1}(p)$	
$\bar{v}_{X2}(p)$	X_2	Z_{21M}	Z_{22M}	Z_{20M}	$Z_{21M} + Z_{X2X1}$	$Z_{22M} + Z_{X2}$	$Z_{20M} + Z_{X2X0}$	$\bar{i}_{X2}(p)$	
$\bar{v}_{X0}(p)$	X_0	Z_{01M}	Z_{02M}	Z_{00M}	$Z_{01M} + Z_{X0X1}$	$Z_{02M} + Z_{X0X2}$	$Z_{00M} + Z_{X0}$	$\bar{i}_{X0}(p)$	

(The left-hand column vector equals the matrix times the right-hand column vector.)

This equation may be rewritten, grouping together like sequence components of current, as

		1	X_1	2	X_2	0	X_0		
$\bar{V}_1(p)$	1	$Z_{11M} + Z_{11L} + Z_{Y1}$	Z_{11M}	$Z_{12M} + Z_{12L} + Z_{Y1Y2}$	Z_{12M}	$Z_{10M} + Z_{10L} + Z_{Y1Y0}$	Z_{10M}	$\bar{I}_1(p)$	
$\bar{v}_{X1}(p)$	X_1	Z_{11M}	$Z_{11M} + Z_{X1}$	Z_{12M}	$Z_{12M} + Z_{X1X2}$	Z_{10M}	$Z_{10M} + Z_{X1X0}$	$\bar{i}_{X1}(p)$	
$\bar{V}_2(p)$	2	$Z_{21M} + Z_{21L} + Z_{Y2Y1}$	Z_{21M}	$Z_{22M} + Z_{22L} + Z_{Y2}$	Z_{22M}	$Z_{20M} + Z_{20L} + Z_{Y2Y0}$	Z_{20M}	$\bar{I}_2(p)$	(24)
$\bar{v}_{X2}(p)$	X_2	Z_{21M}	$Z_{21M} + Z_{X2X1}$	Z_{22M}	$Z_{22M} + Z_{X2}$	Z_{20M}	$Z_{20M} + Z_{X2X0}$	$\bar{i}_{X2}(p)$	
$\bar{V}_0(p)$	0	$Z_{01M} + Z_{01L} + Z_{Y0Y1}$	Z_{01M}	$Z_{02M} + Z_{02L} + Z_{Y0Y2}$	Z_{02M}	$Z_{00M} + Z_{00L} + Z_{Y0}$	Z_{00M}	$\bar{I}_0(p)$	
$\bar{v}_{X0}(p)$	X_0	Z_{01M}	$Z_{01M} + Z_{X0X1}$	Z_{02M}	$Z_{02M} + Z_{X0X2}$	Z_{00M}	$Z_{00M} + Z_{X0}$	$\bar{i}_{X0}(p)$	

The symmetrical-component voltage column vector contains the active armature e.m.f.'s of the synchronous machine (in combined symmetrical-component form), and it is thus possible to solve eqn. (24) for the sequence currents.

recovery voltage transient is evaluated by superimposing these calculated voltages on the voltage existing at the circuit-breaker location prior to fault clearance.

A 3-phase-to-earth fault will be applied to the system of Fig. 3,
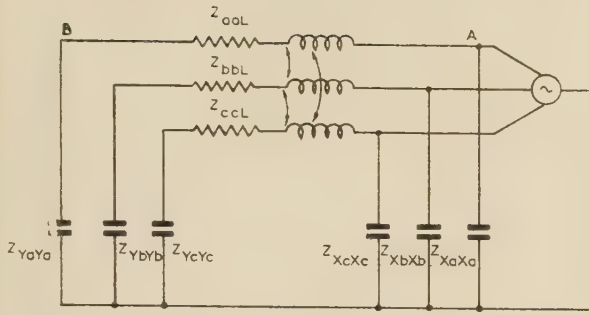
Fig. 3.—Three-phase power system operating normally.

and the behaviour of the system investigated during the period of fault clearance as the contacts of the synchronous machine circuit-breaker open sequentially.

(3.1) Application of Three-Phase-to-Earth Fault at the Remote End of Line

The conditions existing when a 3-phase-to-earth fault has been applied to the above system at B may be regarded as the primitive state for the succeeding analysis. Although a rigorous equivalent-circuit representation is not possible, the balanced fault condition may be symbolized by Fig. 4.

The symmetrical-component operational voltage matrix equation for this balanced fault state is consequently
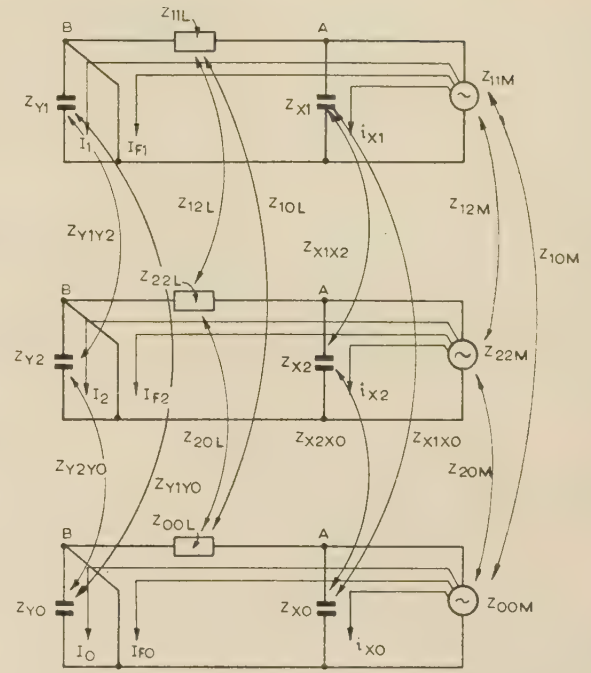
Fig. 4.—Symbolic representation of 3-phase system with 3-phase fault applied at B.

		1	X_1	F_1	2	X_2	F_2	0	X_0	F_0		
$\bar{V}_1(p)$	1	$Z_{11M}+Z_{11L}+Z_{Y1}$	Z_{11M}	$Z_{11M}+Z_{11L}$	$Z_{12M}+Z_{12L}+Z_{Y1Y2}$	Z_{12M}	$Z_{12M}+Z_{12L}$	$Z_{10M}+Z_{10L}+Z_{Y1Y0}$	Z_{10M}	$Z_{10M}+Z_{10L}$		$\bar{I}_1(p)$
$\bar{v}_{X1}(p)$	X_1	Z_{11M}	$Z_{11M}+Z_{X1}$	Z_{11M}	Z_{12M}	$Z_{12M}+Z_{X1X2}$	Z_{12M}	Z_{10M}	$Z_{10M}+Z_{X1X0}$	Z_{10M}		$\bar{i}_{X1}(p)$
$\bar{V}_{F1}(p)$	F_1	$Z_{11M}+Z_{11L}$	Z_{11M}	$Z_{11M}+Z_{11L}$	$Z_{12M}+Z_{12L}$	Z_{12M}	$Z_{12M}+Z_{12L}$	$Z_{10M}+Z_{10L}$	Z_{10M}	$Z_{10M}+Z_{10L}$		$\bar{I}_{F1}(p)$
$\bar{V}_2(p)$	2	$Z_{21M}+Z_{21L}+Z_{Y2Y1}$	Z_{21M}	$Z_{21M}+Z_{21L}$	$Z_{22M}+Z_{22L}+Z_{Y2}$	Z_{22M}	$Z_{22M}+Z_{22L}$	$Z_{20M}+Z_{20L}+Z_{Y2Y0}$	Z_{20M}	$Z_{20M}+Z_{20L}$		$\bar{I}_2(p)$
$\bar{v}_{X2}(p)$	X_2	Z_{21M}	$Z_{21M}+Z_{X2X1}$	Z_{21M}	Z_{22M}	$Z_{22M}+Z_{X2}$	Z_{22M}	Z_{20M}	$Z_{20M}+Z_{X2X0}$	Z_{20M}	$=$	$\bar{i}_{X2}(p)$
$\bar{V}_{F2}(p)$	F_2	$Z_{21M}+Z_{21L}$	Z_{21M}	$Z_{21M}+Z_{21L}$	$Z_{22M}+Z_{22L}$	Z_{22M}	$Z_{22M}+Z_{22L}$	$Z_{20M}+Z_{20L}$	Z_{20M}	$Z_{20M}+Z_{20L}$		$\bar{I}_{F2}(p)$
$\bar{V}_0(p)$	0	$Z_{01M}+Z_{01L}+Z_{Y0Y1}$	Z_{01M}	$Z_{01M}+Z_{01L}$	$Z_{02M}+Z_{02L}+Z_{Y0Y2}$	Z_{02M}	$Z_{02M}+Z_{02L}$	$Z_{00M}+Z_{00L}+Z_{Y0}$	Z_{00M}	$Z_{00M}+Z_{00L}$		$\bar{I}_0(p)$
$\bar{v}_{X0}(p)$	X_0	Z_{01M}	$Z_{01M}+Z_{X0X1}$	Z_{01M}	Z_{02M}	$Z_{02M}+Z_{X0X2}$	Z_{02M}	Z_{00M}	$Z_{00M}+Z_{X0}$	Z_{00M}		$\bar{i}_{X0}(p)$
$\bar{V}_{F0}(p)$	F_0	$Z_{01M}+Z_{01L}$	Z_{01M}	$Z_{01M}+Z_{01L}$	$Z_{02M}+Z_{02L}$	Z_{02M}	$Z_{02M}+Z_{02L}$	$Z_{00M}+Z_{00L}$	Z_{00M}	$Z_{00M}+Z_{00L}$		$\bar{I}_{F0}(p)$

$$\ldots\ldots\quad(25)$$

$$[\bar{V}(p)]_1 = [Z(p)]_1[I(p)]_1 \quad\ldots\ldots\ldots\ldots\ldots\quad(26)$$

(3.2) Sequential Clearance of Circuit-Breaker Phases

The initial conditions prior to fault clearance will be considered to be the steady-state solution of eqn. (25), i.e. the 3-phase-to-earth fault condition will be assumed to have existed for a period long enough for transient effects associated with its application to have disappeared, from time $t = 0$ to $t = t_1$ seconds.

(3.2.1) Clearance of First Phase.

Considering phase 'a' to be the first phase to clear, the constraints imposed at B are those of an open conductor on phase 'a' and a double line-to-earth fault on phases 'b' and 'c'.

Thus, the open-conductor constraint can be imposed by constraining the series-impedance currents I_1, I_2 and I_0 according to eqn. (27), namely

$$I_1 + I_2 + I_0 = 0 \quad . \quad . \quad . \quad . \quad . \quad (27)$$

and the double line-to-earth fault condition is imposed by constraining the I_F mesh currents, namely

$$I_{F0} = -(I_{F1} + I_{F2}) \quad . \quad . \quad . \quad . \quad (28)$$

The current-constraint matrix may now be built up, and is given as

$$[C_2] =$$

	1	X_1	F_1	2	X_2	F_2	X_0
1	1						
X_1		1					
F_1			1				
2				1			
X_2					1		
F_2						1	
0	−1			−1			
X_0							1
F_0			−1			−1	

$$. \quad (29)$$

The operational voltage matrix equation defining this condition is therefore

$$[\bar{V}(p)]_2 = [Z(p)]_2 \cdot [\bar{I}(p)]_2 \quad . \quad . \quad . \quad . \quad (30)$$

where

$$[\bar{V}(p)]_2 = [\bar{C}_{t2}] \cdot [\bar{V}(p)]_1 \quad . \quad . \quad . \quad . \quad (31)$$

and

$$[Z(p)]_2 = [\bar{C}_{t2}] \cdot [Z(p)]_1 \cdot [C_2] \quad . \quad . \quad . \quad (32)$$

The intermediate matrix product $[\bar{C}_{t2}] \cdot [Z(p)]_1$ is given in eqn. (33) below:

$$[\bar{C}_{t2}] \cdot [Z(p)]_1 =$$

	1	X_1	F_1	2	X_2	F_2	0	X_0	F_0
1	$Z_{11M}+Z_{11L}+Z_{Y1}-Z_{01M}-Z_{01L}-Z_{Y0Y1}$	$Z_{11M}-Z_{01M}$	$Z_{11M}+Z_{11L}-Z_{01M}-Z_{01L}$	$Z_{12M}+Z_{12L}+Z_{Y1Y2}-Z_{02M}-Z_{02L}-Z_{Y0Y2}$	$Z_{12M}-Z_{02M}$	$Z_{12M}+Z_{12L}-Z_{02M}-Z_{02L}$	$Z_{10M}+Z_{10L}+Z_{Y1Y0}-Z_{00M}-Z_{00L}-Z_{Y0}$	$Z_{10M}-Z_{00M}$	$Z_{10M}+Z_{10L}-Z_{00M}-Z_{00L}$
X_1	Z_{11M}	$Z_{11M}+Z_{X1}$	Z_{11M}	Z_{12M}	$Z_{12M}+Z_{X1X2}$	Z_{12M}	Z_{10M}	$Z_{10M}+Z_{X1X0}$	Z_{10M}
F_1	$Z_{11M}+Z_{11L}-Z_{01M}-Z_{01L}$	$Z_{11M}-Z_{01M}$	$Z_{11M}+Z_{11L}-Z_{01M}-Z_{01L}$	$Z_{12M}+Z_{12L}-Z_{02M}-Z_{02L}$	$Z_{12M}-Z_{02M}$	$Z_{12M}+Z_{12L}-Z_{02M}-Z_{02L}$	$Z_{10M}+Z_{10L}-Z_{00M}-Z_{00L}$	$Z_{10M}-Z_{00M}$	$Z_{10M}+Z_{10L}-Z_{00M}-Z_{00L}$
2	$Z_{21M}+Z_{21L}+Z_{Y2Y1}-Z_{01M}-Z_{01L}-Z_{Y0Y1}$	$Z_{21M}-Z_{01M}$	$Z_{21M}+Z_{21L}-Z_{01M}-Z_{01L}$	$Z_{22M}+Z_{22L}+Z_{Y2}-Z_{02M}-Z_{02L}-Z_{Y0Y2}$	$Z_{22M}-Z_{02M}$	$Z_{22M}+Z_{22L}-Z_{02M}-Z_{02L}$	$Z_{20M}+Z_{20L}+Z_{Y2Y0}-Z_{00M}-Z_{00L}-Z_{Y0}$	$Z_{20M}-Z_{00M}$	$Z_{20M}+Z_{20L}-Z_{00M}-Z_{00L}$
X_2	Z_{21M}	$Z_{21M}+Z_{X2X1}$	Z_{21M}	Z_{22M}	$Z_{22M}+Z_{X2}$	Z_{22M}	Z_{20M}	$Z_{20M}+Z_{X2X0}$	Z_{20M}
F_2	$Z_{21M}+Z_{21L}-Z_{01M}-Z_{01L}$	$Z_{21M}-Z_{01M}$	$Z_{21M}+Z_{21L}-Z_{01M}-Z_{01L}$	$Z_{22M}+Z_{22L}-Z_{02M}-Z_{02L}$	$Z_{22M}-Z_{02M}$	$Z_{22M}+Z_{22L}-Z_{02M}-Z_{02L}$	$Z_{20M}+Z_{20L}-Z_{00M}-Z_{00L}$	$Z_{20M}-Z_{00M}$	$Z_{20M}+Z_{20L}-Z_{00M}-Z_{00L}$
X_0	Z_{01M}	$Z_{01M}+Z_{X0X1}$	Z_{01M}	Z_{02M}	$Z_{02M}+Z_{X0X2}$	Z_{02M}	Z_{00M}	$Z_{00M}+Z_{X0}$	Z_{00M}

$$(33)$$

and
$$[\bar{V}(p)]_2 = \begin{array}{c|c} 1 & \bar{V}_1(p) - \bar{V}_0(p) \\ \hline X_1 & \bar{v}_{X1}(p) \\ \hline F_1 & \bar{V}_{F1}(p) - \bar{V}_{F0}(p) \\ \hline 2 & \bar{V}_2(p) - \bar{V}_0(p) \\ \hline X_2 & \bar{v}_{X2}(p) \\ \hline F_2 & \bar{V}_{F2}(p) - \bar{V}_{F0}(p) \\ \hline X_0 & \bar{v}_{X0}(p) \end{array} \qquad . \quad . \quad (34)$$

The final product $[\bar{C}_{t2}] . [Z(p)]_1 . [C_2]$ may be evaluated in a similar manner to form $[Z(p)]_2$.

Eqn. (30) defines the performance of the system for the period during which the current in only one phase of the circuit-breaker has been interrupted, i.e. from $t = t_1$ to $t = t_2$ seconds. At $t = t_2$ seconds, the second phase of the circuit-breaker clears.

In order to derive the transient recovery phase voltage to neutral at the circuit-breaker terminals, it is first necessary to solve eqn. (30) for the sequence currents $[\bar{I}(p)]_2$, from which the corresponding sequence voltages to neutral at B, the circuit-breaker location, can be evaluated.

Having transformed these sequence voltages to phase quantities, the recovery voltage transient may finally be obtained by superimposing these calculated voltages on the voltage existing at B prior to the conditions described in eqn. (30). Since the previous condition was that of a 3-phase-to-earth fault at B, the 'existing' voltage at the instant of clearance of the first phase of the circuit-breaker is zero, and therefore the transient recovery voltage for the interval $t = t_1$ to $t = t_2$ seconds will be solely that transient obtained from eqn. (30) as described above.

(3.2.2) Clearance of Second Phase.

The condition created by the opening of the second phase of the circuit-breaker at B is that of a triple dissymmetry in which phases 'a' and 'b' are open, and phase 'c' is earthed.

The constraints imposed by these unbalances are consequently

$$I_1 = hI_2 = h^2 I_0 . \quad . \quad . \quad . \quad . \quad (35)$$

and

$$I_{F1} = hI_{F2} = h^2 I_{F0} \quad . \quad . \quad . \quad . \quad (36)$$

By Kron,[2] the connection matrix $[C_3]$ must again be formed between those currents of this new condition and those of the initial (primitive) balanced fault condition.

$$[C_3] = \begin{array}{c|ccccc} & X_1 & X_2 & 0 & X_0 & F_0 \\ \hline 1 & & & h^2 & & \\ X_1 & 1 & & & & \\ F_1 & & & & & h^2 \\ 2 & & & h & & \\ X_2 & & 1 & & & \\ F_2 & & & & & h \\ 0 & & & 1 & & \\ X_0 & & & & 1 & \\ F_0 & & & & & 1 \end{array} \quad . \quad (37)$$

The new voltage matrix equation of performance will be

$$[\bar{V}(p)]_3 = [Z(p)]_3 . [\bar{I}(p)]_3 \quad . \quad . \quad . \quad (38)$$

where

$$[\bar{V}(p)]_3 = [\bar{C}_{t3}] . [\bar{V}(p)]_1 \quad . \quad . \quad . \quad (39)$$

and

$$[Z(p)]_3 = [\bar{C}_{t3}] . [Z(p)]_1 . [C_3] \quad . \quad . \quad (40)$$

The impedance matrix $[Z(p)]_3$ is consequently

$[Z(p)]_3 =$

$Z_{11M} + Z_{X1}$	$Z_{12M} + Z_{X1X2}$	$h^2 Z_{11M} + h Z_{12M} + Z_{10M}$	$Z_{10M} + Z_{X1X0}$	$h^2 Z_{11M} + h Z_{12M} + Z_{10M}$
$Z_{21M} + Z_{X2X1}$	$Z_{22M} + Z_{X2}$	$h^2 Z_{21M} + h Z_{22M} + Z_{20M}$	$Z_{20M} + Z_{X2X0}$	$h^2 Z_{21M} + h Z_{22M} + Z_{20M}$
$hZ_{11M} + h^2 Z_{21M} + Z_{01M}$	$hZ_{12M} + h^2 Z_{22M} + Z_{02M}$	$(Z_{11M} + Z_{11L} + Z_{Y1} + Z_{22M}$ $+ Z_{22L} + Z_{Y2} + Z_{00M}$ $+ Z_{00L} + Z_{Y0})$ $+ h(Z_{21M} + Z_{21L} + Z_{Y2Y1}$ $+ Z_{02M} + Z_{02L} + Z_{Y0Y2}$ $+ Z_{10M} + Z_{10L} + Z_{Y1Y0})$ $+ h^2(Z_{01M} + Z_{01L} + Z_{Y0Y1}$ $+ Z_{12M} + Z_{12L} + Z_{Y1Y2}$ $+ Z_{20M} + Z_{20L} + Z_{Y2Y0})$	$hZ_{10M} + h^2 Z_{20M} + Z_{00M}$	$(Z_{11M} + Z_{11L} + Z_{22M} + Z_{22L}$ $+ Z_{00M} + Z_{00L})$ $+ h(Z_{21M} + Z_{21L} + Z_{02M}$ $+ Z_{02L} + Z_{10M} + Z_{10L})$ $+ h^2(Z_{01M} + Z_{01L} + Z_{12M}$ $+ Z_{12L} + Z_{20M} + Z_{20L})$
$Z_{01M} + Z_{X0X1}$	$Z_{02M} + Z_{X0X2}$	$h^2 Z_{01M} + h Z_{02M} + Z_{00M}$	$Z_{00M} + Z_{X0}$	$h^2 Z_{01M} + h Z_{02M} + Z_{00M}$
$hZ_{11M} + h^2 Z_{21M} + Z_{01M}$	$hZ_{12M} + h^2 Z_{22M} + Z_{02M}$	$(Z_{11M} + Z_{11L} + Z_{22M} + Z_{22L}$ $+ Z_{00M} + Z_{00L})$ $+ h^2(Z_{01M} + Z_{01L} + Z_{12M}$ $+ Z_{12L} + Z_{20M} + Z_{20L})$ $+ h(Z_{10M} + Z_{10L} + Z_{21M}$ $+ Z_{21L} + Z_{02M} + Z_{02L})$	$hZ_{10M} + h^2 Z_{20M} + Z_{00M}$	$(Z_{11M} + Z_{11L} + Z_{22M} + Z_{22L}$ $+ Z_{00M} + Z_{00L})$ $+ h(Z_{21M} + Z_{21L} + Z_{02M}$ $+ Z_{02L} + Z_{10M} + Z_{10L})$ $+ h^2(Z_{01M} + Z_{01L} + Z_{12M}$ $+ Z_{12L} + Z_{20M} + Z_{20L})$

$$. \quad . \quad . \quad . \quad (41)$$

The corresponding voltage matrix $[\bar{V}(p)]_3$ is:

$$[\bar{V}(p)]_3 = [\bar{C}_{t3}] \cdot [\bar{V}(p)]_1 = \begin{array}{|c|} \hline \bar{v}_{X1}(p) + \bar{V}_2(p) \\ \hline \bar{v}_{X2}(p) \\ \hline h\bar{V}_1(p) + h^2\bar{V}_2(p) + \bar{V}_0(p) \\ \hline \bar{v}_{X0}(p) \\ \hline h\bar{V}_{F1}(p) + h^2\bar{V}_{F2}(p) + \bar{V}_{F0}(p) \\ \hline \end{array} \quad (42)$$

This condition, defined by eqn. (38), applies from $t = t_2$ to $t = t_3$ seconds, i.e. from the instant of time at which the second phase clears to that at which the current in the third phase is interrupted.

(3.2.3) Clearance of Third Phase.

When the third pole of the circuit-breaker opens, the final conditions imposed are such that

$$I_1 = I_2 = I_0 = 0 \quad . \quad . \quad . \quad . \quad (43)$$

$$I_{F1} = I_{F2} = I_{F0} = 0 \quad . \quad . \quad . \quad (44)$$

The operational voltage matrix equation defining this condition is

$$[\bar{V}(p)]_4 = [Z(p)]_4[I(p)]_4 \quad . \quad . \quad . \quad (45)$$

in which only the I_x currents are acting.

Then

Fig. 6.—Symbolic representation of transmission network with 3-phase fault applied on line P at X.

		X_1	X_2	X_0			
$\bar{v}_{X1}(p)$	X_1	$Z_{11M} + Z_{X1}$	$Z_{12M} + Z_{X1X2}$	$Z_{10M} + Z_{X1X0}$	$\bar{i}_{X1}(p)$		
$\bar{v}_{X2}(p)$	$= X_2$	$Z_{21M} + Z_{X2X1}$	$Z_{22M} + Z_{X2}$	$Z_{20M} + Z_{X2X0}$	$\cdot$	$\bar{i}_{X2}(p)$	
$\bar{v}_{X0}(p)$	X_0	$Z_{01M} + Z_{X0X1}$	$Z_{02M} + Z_{X0X2}$	$Z_{00M} + Z_{X0}$	$\bar{i}_{X0}(p)$		(46)

The transient over-voltages obtained for this final condition will ultimately decay to the steady-state open-circuit values.

The initial representation of the network in the vicinity of the circuit-breaker location must be carried out in close detail in order to ensure accuracy when evaluating the transient solution. In common with the classical current-cancellation method,[9] however, to which this matrix transformation method is related,

Fig. 5.—Double-circuit line with arcing-fault-to-earth on phase 'a' of line P.

it is possible to represent more simply impedances which are electrically remote from the circuit-breaker.

(4) SINGLE-PHASE SWITCHING AND AUTOMATIC RECLOSING

Mutual capacitances existing between phases of a double-circuit overhead line may have a considerable effect upon the magnitudes of transient and steady-state currents and voltages during the clearance of a persistent single-phase arcing-fault-to-earth.

Fig. 5 shows the application of such a fault on one phase of a double-circuit line; it is desired to determine the current which would flow in an arcing-fault-to-earth on one conductor, due to capacitance coupling from the other phases after the faulty phase has been switched out at both ends, and to evaluate the transient phase voltages to earth at one of the circuit-breakers during this switching process.

(4.1) Balanced Fault Condition

The primitive condition considered will be a 3-phase-to-earth fault on line P at X, this being defined by eqn. (47) below, the symbolic representation for this condition being given in Fig. 6.

Equation (47):

The current vector (top row):

$$[\,I_{P1}(p),\ I_{PX1}(p),\ I_{Q1}(p),\ I_{QX1}(p),\ I_{F1}(p),\ I_{L1}(p),\ I_{P2}(p),\ I_{PX2}(p),\ I_{Q2}(p),\ I_{QX2}(p),\ I_{F2}(p),\ I_{L2}(p),\ I_{P0}(p),\ I_{PX0}(p),\ I_{Q0}(p),\ I_{QX0}(p),\ I_{F0}(p),\ I_{L0}(p)\,]$$

	$I_{P1}(p)$	$I_{PX1}(p)$	$I_{Q1}(p)$	$I_{QX1}(p)$	$I_{F1}(p)$	$I_{L1}(p)$	$I_{P2}(p)$	$I_{PX2}(p)$	$I_{Q2}(p)$	$I_{QX2}(p)$	$I_{F2}(p)$	$I_{L2}(p)$	$I_{P0}(p)$	$I_{PX0}(p)$	$I_{Q0}(p)$	$I_{QX0}(p)$	$I_{F0}(p)$	$I_{L0}(p)$
L_0	Z_{P1L0}	Z_{PX1L0}	Z_{Q1L0}	Z_{QX1L0}	Z_{F1L0}	Z_{L1L0}	Z_{P2L0}	Z_{PX2L0}	Z_{Q2L0}	Z_{QX2L0}	Z_{F2L0}	Z_{L2L0}	Z_{P0L0}	Z_{PX0L0}	Z_{Q0L0}	Z_{QX0L0}	Z_{F0L0}	Z_{L0L0}
F_0	Z_{P1F0}																	Z_{L0F0}
Q_{X0}	Z_{P1QX0}																	Z_{L0QX0}
Q_0	Z_{P1Q0}																	Z_{L0Q0}
P_{X0}	Z_{P1PX0}																	Z_{L0PX0}
P_0	Z_{P1P0}																	Z_{L0P0}
L_2	Z_{P1L2}	.																Z_{L0L2}
F_2	Z_{P1F2}	.																Z_{L0F2}
Q_{X2}	Z_{P1QX2}	.	.															Z_{L0QX2}
Q_2	Z_{P1Q2}	.	.															Z_{L0Q2}
P_{X2}	Z_{P1PX2}	.	.	.														Z_{L0PX2}
P_2	Z_{P1P2}	.	.	.														Z_{L0P2}
L_1	Z_{P1L1}	Z_{PX1L1}	.	.	.													Z_{L0L1}
F_1	Z_{P1F1}	Z_{PX1F1}	Z_{Q1F1}	.	.	.												Z_{L0F1}
Q_{X1}	Z_{P1QX1}	Z_{PX1QX1}	Z_{Q1QX1}	Z_{QX1QX1}	.	.												Z_{L0QX1}
Q_1	Z_{P1Q1}	Z_{PX1Q1}	Z_{Q1Q1}	Z_{QX1Q1}	Z_{F1Q1}	.	.											Z_{L0Q1}
P_{X1}	Z_{P1PX1}	Z_{PX1PX1}	Z_{Q1PX1}	Z_{QX1PX1}	Z_{F1PX1}	Z_{L1PX1}	.											Z_{L0PX1}
P_1	Z_{P1P1}	Z_{PX1P1}	Z_{Q1P1}	Z_{QX1P1}	Z_{F1P1}	Z_{L1P1}	Z_{P2P1}	Z_{PX2P1}	Z_{Q2P1}	Z_{QX2P1}	Z_{F2P1}	Z_{L2P1}	Z_{P0P1}	Z_{PX0P1}	Z_{Q0P1}	Z_{QX0P1}	Z_{F0P1}	Z_{L0P1}
$=$	P_1	P_{X1}	Q_1	Q_{X1}	F_1	L_1	P_2	P_{X2}	Q_2	Q_{X2}	F_2	L_2	P_0	P_{X0}	Q_0	Q_{X0}	F_0	L_0

The voltage vector (bottom):

$$[\,V_{P1}(p),\ V_{PX1}(p),\ V_{Q1}(p),\ V_{QX1}(p),\ V_{F1}(p),\ V_{L1}(p),\ V_{P2}(p),\ V_{PX2}(p),\ V_{Q2}(p),\ V_{QX2}(p),\ V_{F2}(p),\ V_{L2}(p),\ V_{P0}(p),\ V_{PX0}(p),\ V_{Q0}(p),\ V_{QX0}(p),\ V_{F0}(p),\ V_{L0}(p)\,]$$

(47)

or $\qquad [\bar{V}(p)]_1 = [Z(p)]_1 \cdot [I(p)]_1$ . . . . (48)

where $Z_{P1P1} = Z_{11M} + Z_{P11L} + Z_{PY1}$

$\qquad\quad Z_{P1PX1} = Z_{11M}$

$\qquad\quad Z_{P1Q1} = Z_{11M} + Z_{PQ11L} + Z_{PQY1}$, etc.

(4.2) Unbalanced Fault Condition

The application of the single-phase arcing-fault-to-earth on line P at X imposes the constraint

$$I_{F1} = I_{F2} = I_{F0} \quad . \quad . \quad . \quad . \quad (49)$$

upon the I_F mesh currents. The connection matrix $[C_1]$ is consequently

$[C_1] =$

	P_1	P_{X1}	Q_1	Q_{X1}	L_1	P_2	P_{X2}	Q_2	Q_{X2}	L_2	P_0	P_{X0}	Q_0	Q_{X0}	F_0	L_0
P_1	1															
P_{X1}		1														
Q_1			1													
Q_{X1}				1												
F_1															1	
L_1					1											
P_2						1										
P_{X2}							1									
Q_2								1								
Q_{X2}									1							
F_2															1	
L_2										1						
P_0											1					
P_{X0}												1				
Q_0													1			
Q_{X0}														1		
F_0															1	
L_0																1

. (50)

whence, by Kron, the unbalanced fault condition may be expressed by

$$[\bar{V}(p)]_2 = [Z(p)]_2 \cdot [I(p)]_2 \quad . \quad . \quad . \quad . \quad (51)$$

where $\qquad [\bar{V}(p)]_2 = [\bar{C}_{t1}] \cdot [\bar{V}(p)]_1 \quad . \quad . \quad . \quad . \quad (52)$

and $\qquad [Z(p)]_2 = [\bar{C}_{t1}] \cdot [Z(p)]_1 \cdot [C_1] \quad . \quad . \quad . \quad (53)$

Consequently,

	P_1	P_{X1}	Q_1	Q_{X1}	L_1	P_2	P_{X2}	Q_2	Q_{X2}	L_2	P_0	P_{X0}	Q_0	Q_{X0}	F_0	L_0	
$V_{P1}(p)$	P_1	Z_{P1P1}	Z_{P1PX1}	Z_{P1Q1}	Z_{P1QX1}	Z_{P1L1}	Z_{P1P2}	Z_{P1PX}	Z_{P1Q2}	Z_{P1QX2}	Z_{P1L2}	Z_{P1P0}	Z_{P1X0}	Z_{P1Q0}	Z_{P1QX0}	Z_{P1F0}	Z_{P1L0}	$I_{P1}(p)$
$V_{PX1}(p)$	P_{X1}	Z_{PX1P1}	Z_{PX1PX1}	Z_{PX1Q1}	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$							Z_{PX1L0}	$I_{PX1}(p)$
$V_{Q1}(p)$	Q_1	Z_{Q1P1}	Z_{Q1PX1}	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$								Z_{Q1L0}	$I_{Q1}(p)$
$V_{QX1}(p)$	Q_{X1}	Z_{QX1P1}	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$								Z_{QX1L0}	$I_{QX1}(p)$
$V_{L1}(p)$	L_1	Z_{L1P1}	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$										Z_{L1L0}	$I_{L1}(p)$
$V_{P2}(p)$	P_2	Z_{P2P1}			$\cdot$	$\cdot$	$\cdot$										Z_{P2L0}	$I_{P2}(p)$
$V_{PX2}(p)$	P_{X2}	Z_{PX2P1}															Z_{PX2L0}	$I_{PX2}(p)$
$V_{Q2}(p)$	Q_2	Z_{Q2P1}															Z_{Q2L0}	$I_{Q2}(p)$
$V_{QX2}(p)$	Q_{X2}	Z_{QX2P1}															Z_{QX2L0}	$I_{QX2}(p)$
$V_{L2}(p)$	L_2	Z_{L2P1}															Z_{L2L0}	$I_{L2}(p)$
$V_{P0}(p)$	P_0	Z_{P0P1}															Z_{P0L0}	$I_{P0}(p)$
$V_{PX0}(p)$	P_{X0}	Z_{PX0P1}															Z_{PX0L0}	$I_{PX0}(p)$
$V_{Q0}(p)$	Q_0	Z_{Q0P1}															Z_{Q0L0}	$I_{Q0}(p)$
$V_{QX0}(p)$	Q_{X0}	Z_{QX0P1}															Z_{QX0L0}	$I_{QX0}(p)$
$V_{F0}(p)$	F_0	Z_{F0P1}															Z_{F0L0}	$I_{F0}(p)$
$V_{L0}(p)$	L_0	Z_{L0P1}	Z_{L0PX1}	Z_{L0QX0}	Z_{L0QX1}	Z_{L0L1}	Z_{L0P2}	Z_{L0PX2}	Z_{L0Q2}	Z_{L0QX2}	Z_{L0L2}	Z_{L0P0}	Z_{L0PX0}	Z_{L0QX0}	Z_{L0QX0}	Z_{L0F0}	Z_{L0L0}	$I_{L0}(p)$

$$\tag{54}$$

(4.3) Clearance of Earth Fault on Line P

In order to determine the transient voltage conditions as the two circuit-breakers clear phase 'a' at the two ends A and B of line P, the following constraints must be imposed:

$$I_{P1} + I_{P2} + I_{P0} = 0 \quad . \quad . \quad . \quad . \quad (55)$$

$$I_{PX1} + I_{PX2} + I_{PX0} = 0 . \quad . \quad . \quad . \quad (56)$$

$$I_{F1} = I_{F2} = I_{F0} \quad . \quad . \quad . \quad . \quad . \quad (57)$$

The corresponding connection matrix $[C_2]$ is therefore

$[C_2] =$

	Q_1	Q_{X1}	L_1	P_2	P_{X2}	Q_2	Q_{X2}	L_2	P_0	P_{X0}	Q_0	Q_{X0}	F_0	L_0
P_1				-1				-1						
P_{X1}					-1					-1				
Q_1	1													
Q_{X1}		1												
F_1													1	
L_1			1											
P_2				1										
P_{X2}					1									
Q_2						1								
Q_{X2}							1							
F_2													1	
L_2								1						
P_0									1					
P_{X0}										1				
Q_0											1			
Q_{X0}												1		
F_0													1	
L_0														1

$$. \quad . \quad . \quad (58)$$

and the final configuration is defined by the operational voltage matrix equation

$$[\bar{V}(p)]_3 = [Z(p)]_3 . [\bar{I}(p)]_3 \quad . \quad . \quad . \quad . \quad (59)$$

where

$$[\bar{V}(p)]_3 = [\bar{C}_{t2}] . [\bar{V}(p)]_1 \quad . \quad . \quad . \quad . \quad (60)$$

and

$$[Z(p)]_3 = [\bar{C}_{t2}] . [Z(p)]_1 . [C_2] \quad . \quad . \quad . \quad (61)$$

Eqn. (59) is of the form

		Q_1	Q_{X1}	L_1	P_2	P_{X2}	Q_2	Q_{X2}	L_2	P_0	P_{X0}	Q_0	Q_{X0}	F_0	L_0		
$\bar V_{Q1}(p)$	Q_1	Z_{Q1Q1}	Z_{Q1QX1}	Z_{Q1L1}	Z_{Q1P2}	Z_{Q1PX2}	Z_{Q1Q2}	Z_{Q1QX2}	Z_{Q1L2}	Z_{Q1P0}	Z_{Q1PX0}	Z_{Q1Q0}	Z_{Q1QX0}	Z_{Q1F0}	Z_{Q1L0}		$I_{Q1}(p)$
$\bar V_{QX1}(p)$	Q_{X1}	Z_{QX1Q1}	Z_{QX1QX1}	Z_{QX1L1}	Z_{QX1P2}	. . .	. . .	. . .	. . .	. . .					Z_{QX1L0}		$I_{QX1}(p)$
$\bar V_{L1}(p)$	L_1	Z_{L1Q1}	Z_{L1QX1}	. . .	. . .	. . .	. . .	. . .							Z_{L1L0}		$I_{L1}(p)$
$\bar V_{P2}(p)$	P_2	Z_{P2Q1}	. . .	. . .	. . .	. . .									Z_{P2L0}		$I_{P2}(p)$
$\bar V_{PX2}(p)$	P_{X2}	Z_{PX2Q1}	. . .	. . .											Z_{PX2L0}		$I_{PX2}(p)$
$\bar V_{Q2}(p)$	Q_2	Z_{Q2Q1}													Z_{Q2L0}		$I_{Q2}(p)$
$\bar V_{QX2}(p)$	Q_{X2}	Z_{QX2Q1}													Z_{QX2L0}	$\cdot$	$I_{QX2}(p)$
$\bar V_{L2}(p)$	L_2	Z_{L2Q1}													Z_{L2L0}		$I_{L2}(p)$
$\bar V_{P0}(p)$	P_0	Z_{P0Q1}													Z_{P0L0}		$I_{P0}(p)$
$\bar V_{PX0}(p)$	P_{X0}	Z_{PX0Q1}													Z_{PX0L0}		$I_{PX0}(p)$
$\bar V_{Q0}(p)$	Q_0	Z_{Q0Q1}													Z_{Q0L0}		$I_{Q0}(p)$
$\bar V_{QX0}(p)$	Q_{X0}	Z_{QX0Q1}													Z_{QX0L0}		$I_{QX0}(p)$
$\bar V_{F0}(p)$	F_0	Z_{F0Q1}													Z_{F0L0}		$I_{F0}(p)$
$\bar V_{L0}(p)$	L_0	Z_{L0Q1}	Z_{L0QX1}	Z_{L0L1}	Z_{L0P2}	Z_{L0PX2}	Z_{L0Q2}	Z_{L0QX2}	Z_{L1L2}	Z_{L0P0}	Z_{L0PX0}	Z_{L0Q0}	Z_{L0QX0}	Z_{L0F0}	Z_{L0L0}		$I_{L0}(p)$

$$. \quad . \quad . \quad . \quad . \quad . \quad . \quad (62)$$

The transient sequence voltages of the voltage column vector in eqn. (62) may finally be combined to give the transient phase voltages on phases 'b' and 'c' of line P, which remain intact after phase 'a' has been opened. These voltages will ultimately decay to the unbalanced steady-state values.

(5) RESISTANCE SWITCHING

Surges in voltage and current produced by switching transmission-line capacitance current or by switching capacitor banks can be minimized by the application of shunt resistance across the circuit-breaker contacts during the clearance of the main contacts.[6-8] Arc restrikes may still occur, however, and it is possible to assess the effect of such restriking upon the transient recovery voltage analytically within certain limits.

The performance of the successive network configurations which exist as the circuit-breaker opens is deduced by solving the differential equations relating voltage and current for each circuit condition. The boundary conditions associated with each of these configurations must also be defined before commencing evaluation.

The equivalent positive-sequence network of a 3-phase unloaded line supplied by a source of e.m.f. is represented in simplified form by Fig. 7. The line has been simulated by a shunt capacitance, and all mutual impedance effects have been ignored. During the interruption of the line charging current, all three phases of the circuit-breaker will be considered to open simultaneously.

Prior to interruption, the equivalent positive-sequence network is that of Fig. 7, the operational voltage equation of performance being

$$(pL + R_0 + 1/pC)i = V \quad . \quad . \quad . \quad . \quad (63)$$

When the main arc is ruptured at the first current zero, the resistance R remains connected across the open contacts, as shown in Fig. 8(a), and the modified equation of performance becomes that of eqn. (64), namely

$$(pL + R_0 + R + 1/pC)i = V \quad . \quad . \quad . \quad . \quad (64)$$
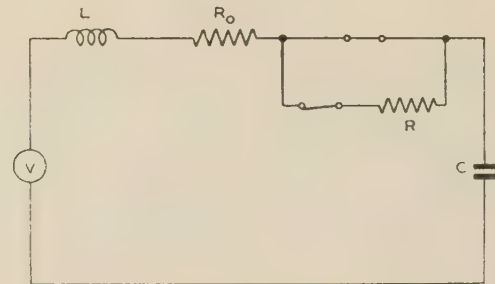
Fig. 7.—Equivalent single-phase circuit of 3-phase system during normal operation.

Depending upon the value of the shunt resistor R, the rate of voltage build-up across the circuit-breaker contacts may exceed the growth of electric strength in the gap, and restriking of the arc may occur. At this instant, the configuration of the positive-sequence network will become that of Fig. 8(b).

(From field tests, the time of incidence of the first restrike will be known to within fairly close limits for a given type of circuit-breaker, and typical system network configuration.)

Arc rupture again occurs at the next power-frequency current zero, or, less probably, at one of the current zeros of the high-frequency component, when the equivalent network configuration will revert to that of Fig. 8(a).

If this second rupture is considered to be the final one, the next change of network configuration occurs when the resistance contacts of the circuit-breaker separate, and becomes that shown in Fig. 8(c). The contact separation time is a linear function of the circuit-breaker contact travel velocity.

If an automatic computing machine is employed for the evaluation of the above calculation, it is possible to obtain multiple solutions by incorporating slight additions to the computer programme of instructions. Several values of the time of first restrike may therefore be considered (within limits ascertained experimentally by field tests), and the most onerous transient over-voltage condition deduced. This facility of obtaining multiple solutions from the computer may also be exploited
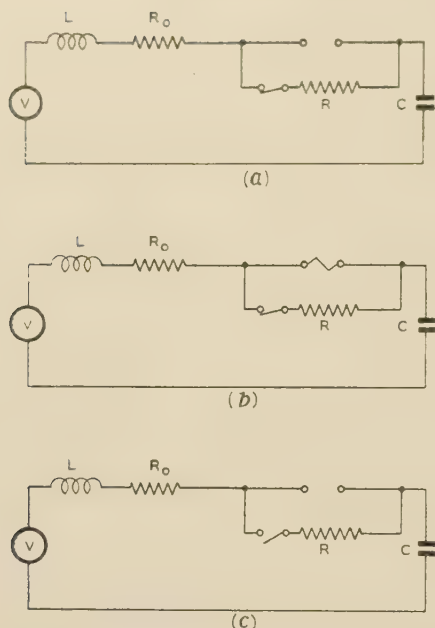
Fig. 8.—Equivalent single-phase circuit of 3-phase transmission system.

    (a) At arc rupture.
    (b) During arc restrike.
    (c) After final interruption.

when assessing the optimum value of circuit-breaker contact shunt resistance to minimize the adverse effects of arc restriking.

Fig. 9 shows the results of such a study. The digital-computer evaluation of transient phase voltage at the transmission line terminals during interruption of the line charging current is compared with the oscillogram of the corresponding field test. The discrepancies shown in the analytical solution were due, in part, to an inaccurate assessment of the instant at which the

first restrike occurred. Further errors were introduced by over-simplification of the positive-sequence equivalent circuit employed, notably by the use of approximate damping factors in the generator-transformer system feeding the overhead line.

(6) CONCLUSIONS

The switching operations in the 3-phase systems considered illustrate the conciseness with which the sequential system unbalances existing during the transient period may be defined using matrix methods. The application of these techniques to larger multi-node networks therefore presents no fundamental complication, in contrast to lattice methods of solution where multiple reflections from numerous nodes incur great complexity. Furthermore, repeated evaluations may readily be obtained by employing automatic computing machines, thereby facilitating the derivation of optimum values of important circuit parameters.

(7) ACKNOWLEDGMENTS

The author wishes to express his appreciation of the encouragement he has received from Dr. W. E. Lewis of the Technical College, Swansea, and Messrs. P. W. Cash and P. B. Jones of the Central Electricity Authority. Acknowledgment is also made to the Central Electricity Authority for permission to publish the paper.

(8) REFERENCES

(1) CHING, Y. K., and ADKINS, B.: 'Transient Theory of Synchronous Generators under Unbalanced Conditions', *Proceedings I.E.E.*, Monograph No. 85 S, December, 1953 (**101**, Part IV, p. 166).
(2) KRON, G.: 'Tensor Analysis of Networks' (Wiley, New York, 1939).
(3) GILL, S.: 'A Process for the Step-by-Step Integration of Differential Equations in an Automatic Digital Computing Machine', *Proceedings of the Cambridge Philosophical Society*, 1951, **47**, Part I, p. 96.
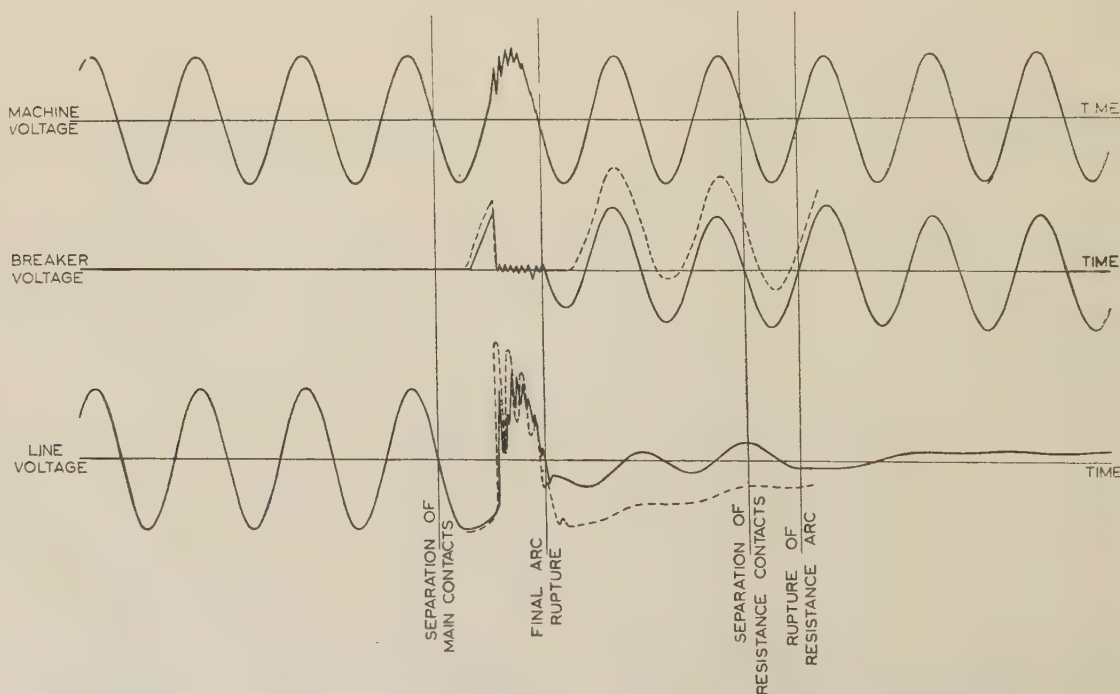
Fig. 9.—Transient voltage characteristics during the interruption of transmission-line charging current.
—————— Field-test characteristics.
– – – – – Digital-computer evaluation.

(4) MORTLOCK, J. R., and HUMPHREY DAVIES, M. W.: 'Power System Analysis' (Chapman and Hall, 1952).

(5) STIGANT, S. A.: 'The Direct Setting-Up of $Z_{\alpha\beta}$ for Closed Mesh Networks from the Network Diagram', *Beama Journal*, 1947, **54**, pp. 28–36 and 65–69.

(6) COX, H. E., and WILCOX, T. W.: 'Influence of Resistance Switching on the Design of H.V. Air-Blast Circuit-Breakers', *Journal I.E.E.*, 1944, **91**, Part II, p. 483.

(7) COX, H. E., and WILCOX, T. W.: 'Performance of H.V. Oil Circuit-Breakers incorporating Resistance Switching', *ibid.*, 1947, **94**, Part II, p. 351.

(8) LEEDS, W. M., and CUSHING, G. B.: 'High-Voltage Oil Circuit-Breakers with Resistance-Equipped Interruptors for Capacitor Switching', *Transactions of the American I.E.E.*, 1954, **73**, Part IIIB, p. 1032.

(9) GOSLAND, L.: 'Restriking Voltage Characteristics in a Power Distribution Network', E.R.A. Report No. G/T86, 1937.

WRITTEN DISCUSSION ON
'ON THE AMPLIFICATION FACTOR OF THE TRIODE'*

Dr. D. A. Bell: It is always difficult to change one's point of view on fundamentals, so that, although Prof. Moullin's argument is right, because it has proved the important point that the 'lumped voltage' treatment of a triode is still valid in the presence of space charge, I find difficulty in absorbing some of the details of the argument. In Section 2, for example, he says 'the field of the anode could not penetrate it [the grid] merely because no anode field existed to do the penetration', and a little lower he speaks of an allegedly fallacious concept of *Durchgriff* which is commonly used 'notwithstanding that no lines of force arise from the charge on the anode'.

One can, of course, say that, since there is zero field within the metallic anode but the integral of flux leaving the charge on the anode equals the charge in question, there must be flux of electric induction, lines of force, in the space inside the anode. But does this flux arise from the charge on the anode? Are we to take it that 'arise from' means 'is caused by', in addition to the geometric sense of the lines either springing from or terminating on the charge? Since experimentally there are no lines, let alone direction arrows on the lines, one cannot say whether they start or finish on a particular charge, and this is probably the root of the representational difficulty. We are, unfortunately, accustomed to think of 'isolated' charges as 'sources' of electric flux, the termination or 'sinks' of the flux being unspecified. But we ought to remember that a line of force always runs between two charges. It is associated with both charges, but it is invidious to suggest that it is due to either of them. To counter the argument that the electric force inside an empty charged conductor is zero, I say that the anode of a diode is not an empty cylinder, and the distribution of charge and field is not the same as if the cathode were absent. The overall boundary conditions at the surfaces of all the electrodes must satisfy the laws of electrostatics, but I am not willing to regard the charge on one of the pairs as solely responsible for the field between the two. Certainly this leads to anomalies in the comparison of the field between a pair of charged parallel plates and the field very near the surface of a finite conductor when the position of the associated 'earth' is not specified. It is at least true of the cylindrical diode that the presence of the anode, with a charge determined by its potential relative to the cathode, affects density of charge on the surface of the cathode (or in the neighbourhood of the potential minimum). According to the traditional concept of field, is this not evidence of a field 'from' the anode at the cathode?

Lt.-Col. A. Krzyczkowski: Prof. Moullin's description of valve physics apparently contradicts those of other authors, and there is a confusion of thought which can be clarified by the following reasoning.

The moment when the anode potential of a thermionic device, with its cathode hot, is suddenly applied should be taken into account. When observing the initiation of the anode current one should expect that some transient period must occur in order to cope with the build-up procedure of such a current. During this period, and especially at its very beginning, the description of physical behaviour by the electrostatic approach seems to be more adequate, as initial currents are very small and their influence on the classical treatment of electric force distribu-

tions is negligible. The anode current builds up its final value during this period owing to the appearance of an anode-cathode potential difference. It is only because of this that the unidirectional movement of electrons within the space between the electrodes becomes possible. One should not forget, therefore, that at the beginning some 'electromotive' force has been required.

Such a force is applied to the thermionic device at the very moment of switching on. Thus it is reasonable to think that the classical picture of electrostatic field penetration from the anode to the space-charge electrons is correct during the short period of the building up of the anode current. After this, the existence between the electrodes of a cloud of rapidly moving electrified particles cannot be ignored, and the only reasonable approach to further analysis is a dynamic one, based on the fact that there is an established current flow within the space between the electrodes. From this point of view the positive charge on the anode has only one purpose for which it should be continuously replenished, i.e. it should neutralize the negative charges delivered by rapidly arriving electrons.

The mechanism of electron movement within the anode-cathode space is as in a self-propelled electronic jet similar to that which can be experimentally observed within a conducting ring, at the absolute zero temperature, and in which an electronic current has been established previously by the same procedure.

Again this picture seems to be correct as long as the potentials of the anode and grid remain unchanged. Once a change in those potentials occurs there must be a transient period, and during this period some sort of superposition of both pictures should take place, as the intimate interplay of static and dynamic approaches seems to be the only logical explanation of the physical behaviour of the device.

Dr. H. Motz: Prof. Moullin takes the equation

$$\ddot{x} = \frac{e}{m\epsilon_0} It \quad \quad \text{. . . . . . (A)}$$

as the starting-point of his treatment of electron motion in a diode. He does not seem to be aware of Llewellyn's equation

$$\ddot{x} = \frac{e}{m\epsilon_0} I \quad \quad \text{. . . . . (B)}$$

where I, the current, is the sum of convection and displacement currents:

$$I = \rho\dot{x} + \epsilon_0 \frac{\partial E}{\partial t} \quad \text{. . . . . (C)}$$

This is the proper basis for a discussion of diodes and leads to the famous Llewellyn–Peterson equations.* The displacement current is ignored in Prof. Moullin's paper. It is seen from eqn. (C) that the field gradient appears in this formula. This gradient is determined by the space charge through

$$\rho = \epsilon_0 \frac{\partial E}{\partial x} \quad \text{. . . . . . (D)}$$

but the potential difference between anode and cathode contributes a constant field in addition to the part of the field supplied by the space charge. The potential difference will be fluctuating

* MOULLIN, E. B.: Monograph No. 211 R, November, 1956 (see **104 C**, p. 222).

* LLEWELLYN, F. B., and PETERSON, L. C.: *Proceedings of the Institute of Radio Engineers*, 1944, **32**, p. 144.

in the a.c. case and will then contribute also to the displacement current. It is clear from eqn. (B) that I is the independent variable for the electron dynamics. This is a feature of any treatment that follows the history of a single electron in the stream. But the rest of the circuit, in turn, determines I.

The suggestion that the electron jet is self-propelling seems misleading. The electrons would not flow without the power provided by the battery. It is like saying that a torrent rushes down a mountain because the back of the water is pushing the front, and not to refer to the gravitational potential.

Prof. Moullin states that electrons cannot reach a negative grid wire: this is misleading. A negatively charged grid wire can collect electrons when its potential is more positive than that of the cathode, even when initial velocity of the electrons is neglected.

The paper argues with various textbooks. These are, however, all very old—dating back essentially to before 1935. A book by K. R. Spangenberg* is more up to date; the account of triode behaviour given there is based on J. H. Fremlin's treatment.† This does not seem to be open to any of Prof. Moullin's objections. It leads to a formula

$$g_m = \frac{3 \cdot 5 \cdot 10^{-6} \left(V_g + \frac{1}{\mu} V_a \right)^{1/2}}{K}$$
$$K = \left[g^{4/3} + \frac{1}{\mu} (d + g)^{4/3} \right]^{3/2} \quad \cdots \quad (E)$$

for the mutual conductance of a plane triode. Fremlin has tested this formula experimentally and found it much superior to Miller's older formula

$$g_m = \frac{3 \cdot 5 \cdot 10^{-6} \left(V_g + \frac{1}{\mu} V_a \right)^{1/2}}{g^2 \left(1 + \frac{1}{\mu} \right)^{3/2}} \quad \cdots \quad (F)$$

Fremlin also gives better formulae for μ, which are important when the grid is close to the cathode. They include forces due to the image of the grid with respect to the cathode plane.

It is disturbing to note that Prof. Moullin's work leads to a formula [his eqn. (21)]

$$g_m = \frac{3 \cdot 5 \cdot 10^{-6} \left(V_g + \frac{1}{\mu} V_a \right)^{1/2}}{\left[g^{4/3} S/g + \frac{1}{\mu} (d + g)^{4/3} \right]^{3/2}} \quad \cdots \quad (G)$$

which differs from (E) by the factor S/g by which $g^{4/3}$ is multiplied. According to the physical insight gained from Fremlin's derivation it seems that K should be the same for two triodes with the same μ regardless of the ratio of grid pitch to grid-cathode distance, unless $S/g \gg 1$. It is therefore gratifying to find that S/g owes its appearance to a slip of arithmetic and should be omitted. It is absent in Prof. Moullin's eqn. (23) for the current in a squirrel-cage triode.

It appears that Prof. Moullin's interesting analysis is equivalent to Fremlin's simpler work, where the derivation of (E) is almost immediate. It is not valid for $S/g \gg 1$ and might lead to new formulae if it were extended to cover such ratios.

Prof. F. G. Heymann (*South Africa*): This is an important paper on the triode, but unfortunately there are some incorrect

* 'Vacuum Tubes' (McGraw-Hill Book Co., 1948).
† 'Calculation of Triode Constants', *Philosophical Magazine*, 1939, 27 (7th Series), 709.

statements and unnecessary complications which detract from its value.

Cylindrical Diode.—Since it is not possible for a positive charge to exist on the inside surface of the anode without the presence of equal negative charge somewhere inside the diode, it does not appear correct to say that the anode charge does not produce any field inside the diode.

By incorrect use of Green's theorem in which the negative charge inside the diode is ignored, Prof. Moullin comes to the conclusion that the charge on the anode has nothing to do with the electric field inside the anode. The same train of thought leads to the erroneous statement that concentric cylinders of infinite radius 'never degenerate into a pair of infinite parallel planes'.

The results in the paper are correct because Gauss's theorem has merely been used to determine the field strength at any point inside the diode after postulating a barrier surface at which the field strength is zero. This leads to the commendable simplification of the equation of motion.

Planar Triode.—As shown in Section 8 of the paper, when space-charge-limited current flows, the total space charge between the barrier surface and the grid is several times smaller than the charge on the grid (unit cross-section) and therefore the space charge as such will not have much effect on the field-strength distribution except near the barrier surface where the charge density is high. The important effect of space charge lies in the fact that it gives rise to increased charge on the grid.

The introduction of two voltages which are constants for a given current forms the major contribution of the paper. V' is the voltage required on the anode for a particular current to flow in the absence of the grid, and V_0 is the voltage required at the grid plane for the same current to flow between the cathode and the grid plane in the absence of the anode.

This leads to the following equations, taken from the paper:

$$V_a = 4\pi Q d + V' \quad \cdots \quad (H)$$
$$V_g = QSB - V_0 \quad \cdots \quad (J)$$

These equations may be used directly to find the amplification factor (in the paper unnecessary complications are introduced). Combining (H) and (J),

$$V_a = (V_g + V_0)\frac{4\pi d}{SB} + V' \quad \cdots \quad (K)$$

or
$$V_a - \mu V_g = \mu V_0 + V' \quad \cdots \quad (L)$$

where
$$\mu = \frac{4\pi d}{SB}$$

and
$$B \simeq -2 \log_e (1 - \varepsilon^{-2\pi c/S}) \quad \cdots \quad (M)$$

The voltages V' and V_0 are constants for a particular current and therefore $(V_a - \mu V_g)$ will be constant at constant current. At cut-off both V' and V_0 are zero.

The cylindrical triode may be dealt with in the same simple manner.

Prof. E. B. Moullin (*in reply*): Dr. Bell's contribution is very appropriate to the paper. But beyond what I wrote there, I fear I cannot say much to dispose of Dr. Bell's difficulties. One can start with the concepts of lines of force and deduce therefrom the concept of electric charge, or vice versa. Personally, I prefer to start with the concept of charge, but that is purely a matter of taste: it is futile to insist that one view is essentially more right than the other. Perhaps it is relevant to remark that it is possible to talk about the pressure of a gas or to describe it in terms of molecular impacts. Either view is admissible; what is not admissible is to use them simultaneously.

Possibly Dr. Bell is subconsciously overlooking the fact that in a diode the charge on the cathode is neutralized by the electron gas between it and the barrier. Hence there would be no field in the space between barrier and anode if there was not space charge in this region. I think he would be prepared to admit that any field in this space arises solely from this space charge and is not affected by the positive charge on the anode surface, which exists only to balance the total space charge. Conceivably we apply slightly different connotations to the term *Durchgriff*. But if this word is interpreted in the way I hoped I had explained in the paper, then surely the statements I made about the word are intelligible and have definite meaning.

Col. Krzyczkowski has raised a very relevant problem in respect of the settling down to the final steady value of current and one which I often pondered upon in the days when I was struggling with 'shot voltage' problems. Consider first a cathode inside a coaxial cylindrical anode, of considerable radius, after the form depicted in Fig. 1 of the paper. Let the cathode be at a temperature such that the total emission is J. It is well known that the anode must be made slightly negative to the cathode before the anode current is reduced to zero; and that by plotting the logarithm of the anode current against this negative potential a straight line is obtained whose slope gives a measure of the cathode temperature. The electric force between cathode and anode must be directed towards the cathode in order to reduce the velocity of some electrons to zero. This means that the charge on the cathode must be positive and its amount in excess of the charge on the electron gas in the space. At the surface of the cylindrical anode, the radial electric force will be proportional to the difference between the positive charge on the cathode and the charge on the electron gas. Since the electric force through the radial thickness of the anode, however thin, must be zero, it thus appears that the charge on the anode must be negative, its sole function being to make the force zero through the thickness of the anode; this charge produces no force in the annular space between anode and cathode. The anode does not repel the electrons: they are attracted to the cathode by the positive charge on it. Surely it now follows that when the anode is given a very slightly positive potential, the charge on it can still remain negative, although at first sight this seems surprising. When a barrier surface (i.e. a surface over which the radial force is zero) exists at a radius smaller than the anode, it follows that the charge on the anode must be positive in order to make the radial force zero through its thickness, by neutralizing the field of the electron gas between the barrier and the anode. Hence it now follows that a condition can obtain when the charge on the anode surface is just zero, in passing from negative to positive charge.

Then the external field of the anode will be zero, since there is no charge on it. Therefore the external field of all the charges (cathode plus electron gas) within the anode must also be zero; this means the positive charge on the cathode is equal to the total charge on the electron gas between cathode and anode. This is equivalent to saying that a barrier surface exists and its radius just coincides with that of the anode. If the cathode is now made very slightly less positive, an equal amount of electron gas will be enclosed in a radius very slightly smaller than the anode; in other words, a barrier surface will exist just inside the anode, which must have positive charge on it, equal to that on the gas between the new barrier and the anode. As the charge on the cathode is decreased still more, the barrier must approach nearer to the cathode. As the positive charge on the cathode is decreased progressively, a state will be reached when the barrier coincides with the cathode; then the total emission will pass to the anode and cannot be further increased.

It appears to me that the transient portion of the anode current must be determined dominantly by the dynamics of the movement of the barrier in response to changes of positive charge on the cathode. It may be essential to take account of the Maxwellian distribution of electron velocities: if this really is necessary, then the analysis, even if it can be made, will inevitably turn out to be too cumbersome to interpret satisfactorily. It might well be illuminating to obtain the solution for a homogeneous stream emitted from the cathode. Possibly some very old work of mine* might be of some assistance in this.

Dr. Motz opens his discussion by saying that I do not seem to be aware of the equation $\ddot{x} = (e/m)I$. In fact I have been aware of this equation for a very long time and used it as the start of my discussion of the paper on 'Transit Time Effects in Diodes' by R. W. Sloane and E. G. James.† At that time this equation was very much in the forefront of my mind, and my impression is that it got there from studying the then recent paper by Mr. W. E. Benham. I recollect the feeling that it must have an obvious direct interpretation which had not then occurred to me, but a few weeks after that it must have done so, because in the same year I published a letter on 'The Three Halves Power Law of the Diode.'‡ In it I started with the equation $\ddot{x} = (e/m)It$ as the direct algebraic statement of the existence of a barrier plane. Since then, 20 years ago, I have regarded the equation involving $\ddot{x}$ as being somewhat *vieux jeux*, whereas Dr. Motz gives 1944 as the date of its emergence. Why he should regard it as the 'proper basis for a discussion of diodes', rather than a direct statement of the existence of a barrier plane, I do not comprehend. He talks of the constant field contributed by the charges on the anode and the cathode. The net charge between barrier and cathode surface is nil: need more be said? In what way is the suggestion that the electron jet is self-propelling misleading? Let us use anthropomorphic terms and ask how an electron in flight can perceive that a battery is connected between anode and cathode. The electron in flight can, so to speak, count up the electrons chasing it and knows that a barrier surface exists at a given constant radius: it knows no more. It realizes being suddenly brought to rest by a collision, and thereby perceives the existence of a metal anode plate, but does not yet perceive there is a battery connected to it. Suppose the battery had been disconnected at some instant after the electron had left the barrier but not yet reached the anode. It would have perceived that the cathode temperature had not changed and also that the positive charge on the cathode was in the process of vanishing. Thus, and thus only, would it have recognized that a battery must be involved in the electrical system: the barrier was catching it up. In the steady state the kinetic energy which was robbed from it on collision with the anode was given it by force arising from the electrons between it and the barrier and not by the battery. What the battery did was somehow to compel it to mingle in a crowd of density about 10^{29} electrons per cm^3, which began to diminish only when it escaped from the cathode surface again and was, once more, slowing down towards the barrier. Just how the battery brought about the crowding up, the electron does not know. The jet in a steam turbine is self-propelling, by means of molecular collisions: there is no need to describe this process by saying that the steam at the back is pushing the steam in front. The steam density at the entrance to the nozzle, and behind it in the boiler, exists because there is a stoker shovelling coal which, at some time, involved the action of the sun. If the stoker ceased to stoke the pressure in the boiler would gradually fade away. There must be a battery to maintain the jet in a state when it is able to be self-propelling.

A negatively charged grid wire could collect electrons emitted

---

* 'The Apparent Inter-Electrode Capacitance of a Planar Diode', *Journal I.E.E.*, 1937, **81**, p. 667.
† *Journal I.E.E.*, 1936, **79**, p. 291.   ‡ *Wireless Engineer*, 1937, **14**, p. 193.

from a cathode which had an even greater negative charge density. (Dr. Motz would say that the potential of the grid, though negative, was positive to the cathode.) I was referring to a system with a barrier surface, in which the charge on the cathode must be positive.

Dr. Motz points out that the expression for g_m derived from my eqn. (21) differs from the classic formula of Dr. J. H. Fremlin because I have made an arithmetical slip in that equation. I thank Dr. Motz for having discovered this slip and for pointing it out: this having been attended to, Dr. Fremlin's method of approach and mine are in agreement in the results they yield.

Turning now to the very welcome discussion from Pretoria, I find it very difficult to phrase my rejoinder to Prof. Heymann. The first portion of his comments, which have had to be abridged in the printed version, are dealt with, in essence, by my replies to Dr. Bell. His initial difficulties might not have arisen if I had more often used the term 'electric force' rather than 'electric field'. It may well be true that I have made an incorrect use of Green's theorem, but as Prof. Heymann does not indicate the precise point where my use was incorrect, I do not see how to correct my mistake. At least in its geometrical sense, surely it is a reasonable statement that 'concentric cylinders of infinite radius never degenerate into a pair of parallel planes': though,

rather remarkably, the two problems do sometimes yield the same arithmetic answer. I always tend to be slightly over-cautious about the use of a one-dimensional configuration in electrical problems; hence my extreme, possibly pedantic, care on this point in the paper. Again, the statement that 'the space charge as such will not have much effect on the field strength distribution except near the barrier . . .' appears to me to demand something further: which is not said. Using a catch-phrase in a purely inquiring sense, may I not be allowed to say 'So what?' Certainly Prof. Heymann arrives at the equation $V_a - \mu V_g = \mu V_0 + V'$ more rapidly than I did. In several places, for example, concentric cylinders of infinite radius, the essence of my paper could have been written in a more telegraphic form. Prolonged discussions with friends, and also the referees, drove me repeatedly to write more and more *in extenso*. Surely the point of such a paper, if indeed it has a point, is not that it should be presented in the shortest way possible for teaching, assuming that it is ever used as a teaching medium. Will not an individual teacher make such shortenings as appeal to his mind, or to that of his class? May it not also be to set out the solution in a way which is likely to contain the answer to almost all the questions which may confront a teacher from his class, if and when members of these classes trip up? which they will do.

DISCUSSION ON

'THE EFFECT OF MAGNETIC SATURATION ON THE D.C. DYNAMIC BRAKING CHARACTERISTICS OF A.C. MOTORS'*

Mr. C. B. Wyatt (*communicated*): I find this Monograph to be of very considerable interest, particularly the approach to the braking-torque/speed curve by using the magnetizing current, I_m, as the independent variable.

I have calculated the braking performance of a hydro-motor using (*a*) the method I would usually employ,† and (*b*) the method tabulated in Table 2. Both methods gave identical results.

The rotor is of the single-cage type, but this presented no difficulty since the parameters R_2 and X_2, which incidentally are relatively high, may be conveniently referred to the primary circuit.

Undoubtedly, Dr. Butler's method is more elegant than (*a*). It is also, I think, less laborious for rotors of the single-cage or slip-ring type, and in these cases I propose to adopt it. Method (*a*) cannot claim to be more than systematic trial and error, but it has I think one merit, and that is its generality.

Dr. Butler's method would seem to present some difficulty in the case of 'skin effect' rotors.

'Skin effect' rotors include all those of the 'deep bar' or 'multi-cage' types but not those of solid steel, since in this latter case R_2 and X_2 are functions of both I_2 and S. In such cases the resulting complexity would, I think, render some other method such as (*a*) more useful.

With regard to Section 3.3, 'Conditions for Maximum Torque', I imagine that the definition of $(dV_0/dI_m)_p$ may present some difficulty unless the usual approximation of two intersecting straight lines of different slope is used (referred to in Section 4).

Incidentally, I found all the equations in Section 9 (Appendices) easy to reproduce with one exception, namely

$$\frac{dI_2^2}{dI_m} = -\frac{2I_m + \dfrac{2X_2I_2^2\left(V_0 - I_m\dfrac{dV_0}{dI_m}\right)}{V_0^2}}{1 + \dfrac{2X_2}{X_m}}$$

in Section 9.2.1.

This must be due to some factor which I have overlooked.

Undoubtedly, the subject of d.c. injection braking, to which Dr. Butler's Monograph is a valuable contribution, is becoming of increasing importance in the application field.

Dr. O. I. Butler (*in reply*): I agree that the expression for dV_0/dI_m presents some difficulty, which is why I recommend, in Section 4.2, using eqns. (7) and (11) in preference. I have used the latter and obtained the curves of Figs. 2 and 3 by tabular computation without any real difficulty.

With regard to 'skin effect' rotors, an American I.E.E. *Transactions* paper is awaiting publication in which I have introduced the slight modification envisaged in Section 3.2. The paper shows that the d.c. braking performance of a double-cage induction motor can be predicted to give good agreement with actual test results, particularly when account is taken of the stray load loss.

The derivation of dI_2^2/dI_m is as follows:
From eqn. (1)

$$I_2^2 = \frac{I_1^2 - I_m^2}{1 + 2X_2/X_m} = \frac{u}{v} \text{ (say)}$$

Hence $\dfrac{dI_2^2}{dI_m} = \dfrac{v(du/dI_m) - u(dv/dI_m)}{v^2}$

$$-\frac{2I_m + I_2^2 2X_2 \dfrac{d(1/X_m)}{dI_m}}{1 + 2X_2/X_m} \qquad . \quad . \quad (A)$$

But $\dfrac{d(1/X_m)}{dI_m} = \dfrac{d(I_m/V_0)}{dI_m} = \dfrac{V_0 - I_m\dfrac{dV_0}{dI_m}}{V_0^2}$

and substitution in (A) provides the value of dI_2^2/dI_m given in Section 9.2.1.

It may be of interest to those engaged in the application field to mention that the accurate method of the Monograph has proved useful in demonstrating the validity of using a simple mathematical relationship for the d.c. braking-torque/speed characteristic in calculating the stopping time and energy dissipation of induction motors.*

* BUTLER, O. I.: 'Stopping Time and Energy Loss of A-C Motors with D-C Braking', *Transactions of the American I.E.E.*, 1947, Paper No. 57–132.

DISCUSSION ON
'THE ELECTRIC STRENGTH OF TRANSFORMER OIL'*

Mr. K. H. Weber and **Mr. H. S. Endicott** (*United States: communicated*): This interesting experimental survey raises some questions concerning both the conditions of test and the conclusions drawn. It would be desirable to know the numerical values of gas and water contents of the oils used—for the determination of which standard procedures are, of course, available. The authors conclude from the results given in Table 1 that there is no dependence of electric strength on time in the range of 10 to 60 sec. In view of this, it is not clear how the effect of electrode materials can be explained by the time dependence of electric strength. The authors also suggest that, in contrast with all previously published tests on insulating oils, their data show little dependence of electric strength on electrode size. At this time, we wish only to comment on the particular question of electrode size. On the basis of our recent paper,[A] it is readily shown that the results of the monograph, in their effect on electrode size, are precisely what one would expect from a consistent area-effect theory.

In our paper, the electric strength of oil is shown to depend on electrode size. A study of 1 600 observations of transformer oil breakdown, using four pairs of uniform-field (Rogowski) brass electrodes established the extremal[B] nature of such breakdown distributions, and yielded an area-effect equation which agrees closely with test results. This equation is

$$E_1 - E_2 = \frac{S_E}{\sigma_N} \log_\varepsilon \frac{A_2}{A_1} \qquad (A)$$

where E is the modal strength and S_E is the standard deviation of N measured breakdown values; σ_N is a function of N only.[B]

It is seen from this equation that the difference in breakdown strength depends, not only on the area ratio of the two sets of electrodes, but also on the magnitude of the standard deviation of the measured values. One of the particularly satisfactory features of this equation is that it contains no arbitrary or adjustable constants. Applying it to the data of Table 3 for the 1·0 and 1·5 in diameter electrodes (using means because modal values are not given):

Measured values $\quad E_1 - E_2 = 668 - 649 = 19\,\text{kV/cm}$
(*Zein El-Dine*
and Tropper)

Value calculated $\quad E_1 - E_2 = \dfrac{S_E}{\sigma_N} \log_\varepsilon \dfrac{A_2}{A_1}$
from eqn. (A)

$$= \frac{29\cdot2}{1\cdot16} \log_\varepsilon (1\cdot5)^2 = 20\cdot4\,\text{kV/cm}$$

Thus the 'slight tendency for the electric strength to increase with decreasing size of the electrodes' is predicted quite closely.

* EL-DINE, M. E. Z., and TROPPER, H.; Monograph No. 135 S, June 1953 (see 103 C, p. 35).

The greater changes reported by others can be explained by the larger dispersion of their results, the larger ratio of areas used, or both.

If breakdown measurements in general have these extremal characteristics, the breakdown strength of a liquid is not uniquely determined unless electrode area is also specified; i.e. according to eqn. (A), electric strength is independent of area only when the standard deviation of the observations is zero. (Practically speaking, this would mean that the variations that occur are solely random errors of measurement, which are described by a normal probability curve.)

Extension of these considerations, for example, to measurements of electric strength with pairs of spheres at different spacings, or with different radii, shows that electric strength for a given liquid will not remain constant for these different conditions—even though the usual 'field uniformity criterion' for spheres at close spacing be satisfied. Discrepancies in the literature for functions of the electric strength of various liquids, ostensibly measured under the same conditions, may be attributable to overlooking the experimental and theoretical requirements of the extremal area effect.

Finally, we believe that, in the absence of complete control over the many factors that may affect oil strength, it would appear that the safest way to make comparison among electrode pairs of different size or configuration would be to test them at the same time, when both are physically present in the same oil and subjected to the same conditions of voltage application.

Drs. M. E. Zein El-Dine and **H. Tropper** (*in reply*): The standard procedure[C] referred to by Messrs. Weber and Endicott for determining the gas and water contents of the oil is not sufficiently accurate for most of the tests reported in the monograph. For example, with regard to the gas content which plays a predominant role in the breakdown of pure oils, the oil for the tests was degassed at an equilibrium pressure of 0·5 mm Hg. From Fig. 11 of Reference C its air content can be estimated to be approximately $6\cdot7 \times 10^{-3}$ per cent per volume, and this is less than the precision claimed for the method, which is $\pm0\cdot2$ part in 10^6 by weight, or $1\cdot5 \times 10^{-2}$ per cent per volume. The accuracy no doubt can be improved, but it was found more satisfactory and convenient in subsequent work with closed testing systems to degas the oil at a pressure of 0·01 mm Hg and then to admit measured amounts of air or other gases at various equilibrium pressures.

There seems to be some confusion about the time independence of the electric strength and the bearing of this on the dependence of the electrode material. A distinction must be made here between measurements made with long-duration voltages and measurements using impulse voltages, since the breakdown mechanisms operative in the two cases will not be the same. The experiments have shown the electric strength to be inde-

pendent of time in the range of 10–60 sec for long-duration (i.e. alternating and direct) voltages, and moreover, for such voltages, a marked dependence on electrode material was found. With ⅓ microsec impulse voltages, on the other hand, no electrode dependence was observed, and it was therefore concluded that whatever electrode-sensitive mechanism was operative for long-duration voltages could not have been present when voltages of microsecond duration were used. In this sense, therefore, the mechanism causing an electrode dependence is assumed to be time dependent (or time consuming), as stated in the monograph.

It is gratifying that the results in the monograph on the electrode-size effect conform with the area-effect theory. This theory is perhaps not sufficiently appreciated in this country, and our correspondents' little homily relevant to this theory is therefore helpful and valuable.

It must be emphasized, however, that the purpose of the work in this laboratory on the electric strength of oil, of which the monograph presents some of the preliminary results, is not to find reliable breakdown data for design purposes. The aim is rather to study systematically the factors which determine the electric strength of oil in order to gain a better understanding of the breakdown mechanism. This involves tests whose results do not always lend themselves to a treatment by the statistical theory of extremal values. For example, interesting light is shed on the mechanism by a study of the first few breakdowns, i.e. during the period when an equilibrium is gradually established between the adsorbed gas on the electrodes and the gas in solution in the oil. The corresponding strength values may either increase (conditioning up) or decrease (conditioning down) with successive breakdowns, and the total variation can be appreciable, so that these values are often discarded for statistical analysis. For impulse voltages, very interesting results have recently been

obtained[D] with a single-shot technique, i.e. by using fresh electrodes for every single breakdown. The breakdown values were of a very high consistency and showed a marked difference from the average values determined from a series of breakdowns. They are not only significant from the point of view of the theory of the electric breakdown, but should have an important bearing on insulation failure in practice, for one would expect such a failure to occur at a voltage corresponding to a first breakdown.

Alas, in the course of the work a gap dependence of the electric strength was found. In view of this correspondence the tests were made with particular care and were repeated at different times. We used direct voltages and plate electrodes having a Rogowski profile with a flat portion of 1 cm diameter. The gap spacings were 2, 4 and 6 mm, and the corresponding electric strengths were 330, 263 and 203 kV/cm. It is hoped to publish these results shortly.

REFERENCES

(A) WEBER, K. H., and ENDICOTT, H. S.: 'Area Effect and its Extremal Basis for Electric Breakdown of Transformer Oil', *Transactions of the American I.E.E.*, 1956, **75**, Part III, p. 371.
(B) GUMBEL, E. J.: 'Statistical Theory of Extreme Values and Some Practical Applications', National Bureau of Standards, Applied Mathematics Series No. 33 (U.S. Government Printing Office, Washington, D.C., 1954).
(C) KAUFMAN, R. B., SHIMANSKI, E. J., and MACFADYEN, K. W.: 'Gas and Moisture Equilibrium in Transformer Oil', *Transactions of the American I.E.E.*, 1955, **74**, Part I, p. 312.
(D) HANCOX, R.: 'Dependence of the Dielectric Breakdown of Transformer Oil on Electrode Conditions', *Nature*, 1956, **178**, p. 1305.

DISCUSSION ON
'LAGUERRE FUNCTIONS: TABLES AND PROPERTIES'*

Mr. K. Kraus (*Czechoslovakia: communicated*): I should be interested to know what the authors mean by the symbol p, which is denoted as Heaviside's operator. If two functions F(x) and F[p] are operationally equivalent, i.e. related by the one-sided Laplace transformation, the meaning of this operator p in Heaviside's conception is different from that which we have in mind in the Laplace transformation. According to the Laplace transformation, p denotes a complex number, but Heaviside takes the operator p as an algebraic number. This is due to the difference between the Heaviside conception of the operational calculus and the operational calculus based on the one-sided Laplace transformation, which may lead to errors in the calculation of expressions of the same form as eqns. (57), (58) and (59). In my opinion it would be better to take p directly as a complex number and to use the operational calculus based on the one-sided Laplace transformation. In this way we can avoid mistakes which are not apparent at first sight and which follow from the pure-mathematical conception of one or other operational calculus.

Messrs. J. W. Head and **W. Proctor Wilson** (*in reply*): There is, in our view, no incongruity in referring to the p used in our paper as the Heaviside operator. In modern operational calculus, placed on a sound mathematical basis by Carson, van der Pol, Wagner and many others, the limitations and necessary modifications of Heaviside's original concepts are well understood.†

Expression (60) in the paper defines the relation between F[p] and F(x) in the sense that

(i) It enables F[p] to be derived from F(x) by evaluating the definite integral.
(ii) F(x) may be derived from F[p] by the solution of (60) regarded as an integral equation.

Alternatively, the inverse expression to (60), well known to Heaviside, is the Bromwich integral

$$F(x) = \frac{1}{2\pi i} \int_{c - i\infty}^{c + i\infty} \frac{F[p]}{p} \, dp$$

in which c is chosen so that the path of the integral lies to the right of all singularities of the integrand.

Having thus defined the way in which F[p] is derived from F(x) or F(x) from F[p], we are at liberty to manipulate F[p] algebraically as we please in order to change it into a linear combination of functions $G_1[p]$, $G_2[p]$ ... for which the corresponding $G_1(x)$, $G_2(x)$ are known, and during this process of manipulation we can regard p as a mere complex number.

To sum up, the sense in which the Heaviside operator p is used in our paper is that adopted by the writers in our References 4 and 5.

INDEX TO VOLUME 104, PART C

1957

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

PART C—MONOGRAPHS, SEPTEMBER 1957

CONTENTS